KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Imaging and Pathology

THE DEVELOPMENT OF A METHODOLOGY TO OPTIMIZE THE PERFORMANCE OF BREAST TOMOSYNTHESIS

Eman SHAHEEN

Jury:

Promoter: Prof. dr. ir. Hilde Bosmans

Co-promoter: Prof. dr. Chantal Van Ongeval

Chair: Prof. dr. Philippe Demaerel

Jury members: Prof. dr. Predrag Bakic

Prof. dr. Ria Bogaerts

Prof. dr. Anders Tingberg

Prof. dr. Mireille Van Goethem

Prof. dr. Erik Van Limbergen

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biomedical Sciences

April 2014

Acknowledgments

"And whoever is careful of (his duty to) Allah, He will make for him an outlet.

And give him sustenance from whence he thinks not".

[Surat Al Talaq (2,3)- The Holy Quraan]

These words were my gold rule throughout the journey of my PhD. I started in this group 6 years ago as a pre-doctoral student in a field slightly different from my background. There were ups and downs, successes and failures, but during this trip I learned that nothing valuable comes easy, you have to work hard, do your best, and believe that you will be rewarded at the right moment. I would like to grasp this opportunity to acknowledge the most prominent contributors to this thesis.

This thesis is respectfully presented to Prof. Rik Torfs, Rector of the KU Leuven, to Prof. Dr. Jan Goffin, dean of the Faculty of Medicine, to the current head of the Radiology department Prof. Dr. Raymond Oyen, to the previous head Prof. Dr. Guy Marchal, to my promoter Prof. Dr. ir. Hilde Bosmans and to my co-promoter Prof. Dr. Chantal Van Ongeval.

I would like to gratefully acknowledge Prof. Dr. Philippe Demaerel and Prof. Dr. Steven Dymarwkowski for chairing the defense and the members of the jury: Prof. Dr. Predrag Bakic, Prof. Dr. Ria Bogaerts, Prof. Dr. Anders Tingberg, Prof. Dr. Mireille Van Goethem and Prof. Dr. Erik Van Limbergen for their valuable comments and guidance that improved the quality of this dissertation.

Prof. Hilde Bosmans was my promoter from the beginning. It was a challenge for both of us due to my different background. She took a risk, she didn't know me and I had to prove myself. She taught me a lot, starting from the basics and combined our knowledge to produce this work. I'm so grateful Hilde for giving me this chance. You were always there for me when I needed your help even in your busiest moments. You gave me time, efforts and constant guidance. I know I'm pushy, and not always easy going but it worked out fine. Letting me take my own decisions within the research topic was a freedom that allowed me to grow up and be ready to be independent. I would like to deeply thank you for everything.

My dear co-promoter Prof. Chantal Van Ongeval, thank you for supporting me and guiding me through all the medical aspects in this work. You believed in my work and encouraged me with your precious insights, comments, feedback. Even when you didn't have the luxury of time, you managed to fit me in. I can only say that I'm lucky to have you as a main contributor in this PhD.

I would like to thank every member of the LUCMFR for your direct and indirect contribution to my work. I enjoyed working with you and already miss you all (in alphabetic order): Annelies Jacobs, Andreas Stratis, Elena Salvagnini, Frederik De Keyzer, Prof. Federica Zanca, Guozhi Zhang, Joke Binst, Joris Nens, Kim Lemmens, Koen Michielsen, Lesley Cockmartin, Nausikaä Geeraert, Nicholas Marshall, Nelis Van Peteghem, Xochitl Lopez Rendon. Annelies jij was mijn buur in het Louvre vanaf het begin. Jij heeft me altijd gesteund om beter te zijn. Ik durfde te praten in het Nederlands door jou. Ik waardeer onze vriendschap. Frederik, the gift of LUCMFR, for the past 2 years you were my second author. I never hesitated to brainstorm with you, ask for help or even chat/complain in the corridors. I'm happy to have had the chance to work and learn from you. I wish you the best of luck with your PhD and now you can break the "zeer". Federica, you were the first to welcome me in the group. It was my pleasure working together. I looked up to you when I felt torn between work, home, and kids. You always helped, supported and advised me whenever I asked for it on both the scientific and personal levels. Thank you and I wish you all the best. I know you will shine wherever you go like always. Guozhi my foreigner friend, I'm glad you are coming back, they need you in the group. Lesley, you were always so kind to help everyone in the group, you supported my work and you are a good friend to me. Nausikaä, it took us sometimes to start talking and knowing each other better, but since then you can't imagine how happy I am to become closer friends. Xochitl, you were my neighbor in the Louvre for the last 2 years. I admit, I can be annoying sometimes, but you bore with me, supported and encouraged me for all the times I was feeling down. You were always so excited with my successes whether small or big, thank you for being there for me, I really appreciate it.

I had the luxury to be part of the Optimam project which did not play only the role of funding but also collaboration with its members: Prof. Kenneth Young and Prof. David Dance, thank you for your constant scientific follow-up, feedback, help and support that always lead to enriching my work. Prof. Kevin Wells, Alaleh Rashidnasab, Premkumar Elangovan, Lucy Warren, Alistair Mackenzie, it was a pleasure working with all of you.

I am indebted to the radiologists who participated in the observer studies and spent their lunch breaks to read my images and provide me with valuable insights to improve my simulations: Dr. Andre Van Steen, Dr. Sandra Postema, Dr. Kirsten Joossens, Dr. Ilse Vervloessem, Dr. Riet D'Hauwe, Dr.

Liesbet Vermeiren and Dr. Anne-Sophie Vliegen. I can't forget Ann Similon for her help in the microcalcification study, her warm smile and words every time we meet.

I would like to thank Prof. Dr. Reinhilde Jacobs for introducing me to the professors of the Medical Imaging Research Center (MIRC) where I had the opportunity to start on the Breast CT project as a pre-doctoral student. I owe my appreciation to Prof. Dr. ir. Paul Suetens, the head of the MIRC, for the scientific atmosphere provided in the Louvre that facilitated a number of collaborations for my work and making friends outside my group: Maarten Depypere, Jenny Ceccarini and Vesselina Ferferieva.

I owe my appreciation to the secretary team of the Radiology Department: Ingrid Fruyt, Linda Meersman, Monika Philips, Katrin Lorent and Cindy De Keyzer. I have worked closely with both Ingrid and Linda. You always took care of my administration work with kindness and no complain. I appreciate all your hard work and trying to fit me in your busy schedules.

Living abroad, away from family, is hard. I would like to express my gratitude to my mini Egypt in Belgium: Mohamed & Enas, Mohamed & Latifa, Karim & Noura, Hamdy & Lamia, Marwa Ali, Yousra Abd Elmoteleb for their warm mutual relations, making my stay even more restful and productive. Sinead, I appreciate your support and our talks during the swimming lessons.

At last and definitely not least, my family whose love and support guided me through this journey. To my parents, I would have never been here without you. You are the best parents ever, my role models professionally and personally, I love you. You raised me to be what I am now, gave me support, courage, love and everything you could give. I hope you are proud of me. There are not enough words to describe how thankful and grateful I am to you. My brother Mohamed, I miss you so much, and thank you for being my big brother who always looked out for his little sister. Omar and Adam, my sweetest caring kids, you drive me crazy but you are the best thing in my life, God bless you.

Medhat my beloved husband, you take the credit for all this. Without your continuous encouragement, love, kindness, baring my crazy moments, taking care of the kids whenever needed, supporting me emotionally, …. I could have never finished this thesis. You know we laughed about it (mother of 2 kids, wife and a PhD student, I must be crazy) but we finally did it. I love you and respect you forever.

Emmy

List of Acronyms

2/3D	Two/three dimensional
a-Se	Amorphous Selenium
AND	Logical AND function
AUC	Area Under the Curve
BCT	Breast Computer Tomography
BIRADS	Breast Imaging-Reporting and Data System
C	Contrast
Ca	Calcium
CaC_2O_4	Calcium Oxalate
CADe	Computer Aided Detection
CADx	Computer Aided Diagnosis
CC	Craniocaudal view
CH_2O	Formaldehyde
CI	Confidence Interval
COR	Centre Of Rotation
CT	Computer Tomography
DBT	Digital Breast Tomosynthesis
DSCV	Digital Spot Compression View
DQE	Detective Quantum Efficiency
FBP	Filtered Back Projection
FFDM	Full Field Digital Mammography
FPF	False Positive Fraction
FT	Fourier Transform
IFT	Inverse Fourier Transform
κ	Kappa statistic
micro-CT	Micro Computer Tomography
MLO	Mediolateral Oblique view
MRI	Magnetic Resonance Imaging
MTF	Modulation Transfer Function
NH	Null Hypothesis
NIST	National Institute of Standards and Technology

NURBS	Non Uniform Rational B-spline
PACS	Picture Archiving and Communications System
PMMA	Polymethyl Methacrylate
PV	Pixel Value
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SD	Standard Deviation
SdNR	Signal difference to Noise Ratio
SF	Scatter Fraction
SID	Source to Detector Distance
SPR	Scatter to Primary Ratio
SVM	Support Vector Machine
TPF	True Positive Fraction

Table of Contents

Chapter 1

General Introduction

Chapter 1

1.1. Breast Tomosynthesis

1.1.1. Rationale for digital tomosynthesis

Tomosynthesis is a tomographic imaging technique acquiring a limited number of projections from a narrow angular range, and combining these projections to reconstruct a quasi-3D image [1]. In 1932, Ziedses des Plantes [2] proposed the theoretical framework for limited angle tomography. In the 1970's and 1980's, ectomography and flashing tomosynthesis were developed, both of which are forms of tomosynthesis [3]. Further developments were challenging due to two main reasons: the residual blur from objects outside of the plane of interest and the absence of digital detectors suitable for acquisitions of multiple projections. With the advent of spiral computer tomography (CT) in 1980s, digital tomosynthesis development was halted [3]. However, with the introduction of flat panel radiographic detectors in the late 1990s, research interest to explore tomosynthesis was back and in 1997, Niklason et al. [4] demonstrated the first tomosynthesis imaging of the breast, using the acquisition geometry and a rotating tube of a mammography system (Fig. 1.1). Tomosynthesis has been applied to several clinical applications such as dental imaging, angiography, imaging of the chest, breast and bones [3]. In this thesis, we focus on digital breast tomosynthesis (DBT).

1.1.2. Technical aspects of digital breast tomosynthesis

Most of the current DBT systems consist of the same components as digital mammography systems with an x-ray tube mounted on an arm that moves either continuously or in a step and shoot mode within a limited angular range taking low dose projection images of a compressed breast. The tube rotates around a point close to or on a detector. These projection images are reconstructed to form tomographic slices. Due to this limited angular range, the resolution in the z direction (i.e., in the direction perpendicular to the x-y plane of the projection images) is limited but the resolution in the x-y plane is often superior to CT.

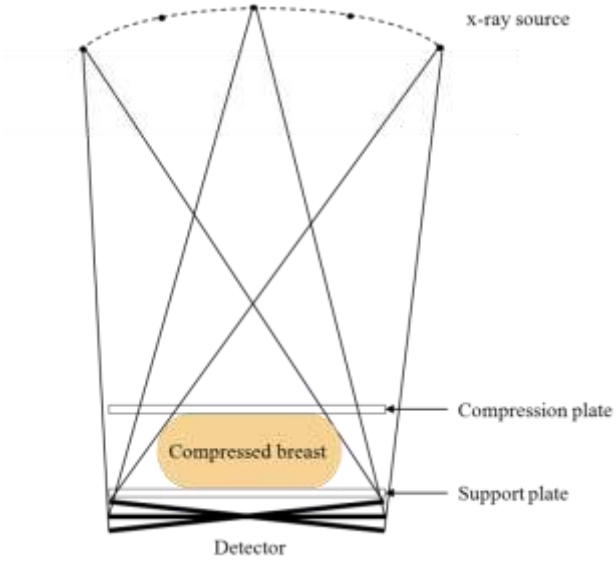


Fig. 1.1 Breast tomosynthesis acquisition geometry

1.1.2.1. Detectors

In addition to the requirements for digital mammography, DBT detectors have special requests to allow acquisition of many projection images during short scan times: 1) faster read out time; 2) minimal ghosting and lag; 3) the best possible detective quantum efficiency (DQE) at low exposures. There are two main detector types for DBT that fulfil these requirements: direct and indirect full field digital detectors. Direct conversion detectors use semiconductors to transform x-ray photons to electric signal with no intermediate step. The most commonly used conversion material is amorphous selenium (a-Se). Indirect conversion detectors first transform x-ray photons to light via a scintillating medium, after which the light photons are converted to electrical charge via photodiode. Another design is based on a scanning slit photon counting detector. The advantages of this design are very low scatter signal, no electronic noise and high quantum efficiency [1].

1.1.2.2. Tube motion

Most available DBT systems operate with a continuous tube motion where x-rays are pulsed during continuous motion of the gantry; inevitably this leads to some blurring due to focal spot motion during each exposure. An alternative approach is the step and shoot method where the gantry comes to a complete stop at each angular location before generating the x-rays. Provided that the tube stops completely (no mechanical jitter or reverberation), there won't be focus motion blurring, although

there may be increased patient motion blurring if the overall scan time is longer than a continuous motion operated over the same angular range with the same number of exposures [5].

1.1.2.3. Acquisition parameters

The optimization of the acquisition parameters including the angular range, scan time, number of projections, etc. is still under investigation. These parameters vary by system design with scan times ranging from 3 to 25 seconds, the angular range varies from 11° to 50° and the number of projections varies from 9 to 25 projection images [1]. There are some investigations [6, 7] for the distribution of the total exposure considering non-uniform distribution, i.e. assigning half of the total exposure to the central projection (0°) and the other half is divided among the remaining projections evenly, instead of the current uniform distribution, i.e. the total exposure is divided among all the projections evenly. The non-uniform distribution design was proposed to combine the strengths of two-dimensional full field digital mammography (2D FFDM) and DBT, by using the central projection image for microcalcification detection and the reconstructed planes for mass detection. The results of the investigations are not yet conclusive.

1.1.2.4. Reconstruction algorithms

Tomosynthesis images are generally reconstructed using filtered back-projection (FBP) or statistical iterative techniques. Due to the limited angular range, tomosynthesis suffers from the residual blur of objects outside the plane of interest. Back-projection reconstructions are inexact and the choice of filters applied to the projection data before the application of back-projection affects the quality of the reconstruction [8]. Alternatively, iterative approaches, such as simultaneous algebraic reconstruction technique and maximum likelihood expectation maximization, provide superior image quality with less artifacts compared to FBP but require greater computation time [9]. The optimization of performance of reconstruction algorithms is an active topic of research [10, 11].

1.1.2.5. Computer aided detection and diagnosis

Computer aided detection (CADe) and computer aided diagnosis (CADx) are sophisticated computer programs designed to recognize patterns in images and to provide assistance in the detection of the presence of disease. They are used in mammography to assist radiologists in the detection of breast lesions, both microcalcifications and masses. The CADe systems are usually confined to marking conspicuous structures and sections. For tomosynthesis, CADe may have the potential to improve detection of subtle lesions, but the current algorithms for 2D mammography cannot be directly

translated to tomosynthesis. The CADx systems evaluate the conspicuous structures. Tomosynthesis CAD algorithms should consider the use of both slice-by-slice evaluation as well as full 3D evaluation. The impact of using CAD is not yet determined.

1.1.2.6. Advanced applications

Breast tomosynthesis may be combined with other modalities such as ultrasound, single-photon emission CT, positron emission tomography, and optical imaging [9]. Some studies report the use of contrast enhanced dual-energy tomosynthesis to provide information similar to contrast enhanced magnetic resonance imaging while keeping the advantages of tomosynthesis in terms of low cost, easy acquisition and high resolution [9].

1.1.3. Clinical aspects

The evaluation performance of a newly introduced modality should match or be better than that of the current modality for all types of lesions. Therefore, the current clinical studies are comparing the performance of DBT to 2D FFDM in terms of detectability and characterization of masses and microcalcifications. The first small scaled studies focused on the detectability of masses. Poplack et al. [12] compared DBT to conventional mammography with 98 cases in a paired side-by-side study where they found that DBT had comparable or superior image quality to that of film-screen mammography. Good et al. [13] compared 2D FFDM to DBT in a paired blinded study of 30 patients (i.e. 2D was evaluated separately from DBT, not as a side-by-side study). They concluded that no significant difference was found for cancer detection between the two modalities. Andersson et al. [14] also compared 2D FFDM to DBT in a population of 40 cancers and concluded that the cancer visibility in DBT was superior to 2D FFDM. Gur et al. [15] investigated the same in a larger sample of 125 patient cases with different set-ups. Their results were consistent with the previous studies and added that the use of DBT may decrease the recall rate by 10%.

The first large prospective study with 513 patients was conducted by Teertstra et al. [16]. They have found that 2D and DBT had the same sensitivity for cancer detection and 2D had higher specificity than DBT. They recommended the use of DBT only as an additional technique to mammography in patients referred with an abnormal screening mammogram or with clinical symptoms. This recommendation was also stated in a recent study by Michell et al. [17].

Furthermore, Gennaro et al. [18] compared the clinical performance of two view 2D FFDM to one view DBT and found that DBT was not inferior to two view 2D. This finding was further confirmed by Wallis et al. [19] and Svahn et al. [20] for other DBT systems. Wallis et al. [19] found no significant difference in diagnostic accuracy in their set-up, but two view DBT outperformed two view 2D mammography only for the less experienced readers.

The use of synthetic 2D images is highly motivated in an attempt to replace the extra exposure of the 2D image. Gur et al. [21] investigated the value of synthetic 2D images combined with a DBT series. The results showed lower sensitivity when using synthetic 2D images instead of real 2D images but comparable specificity.

Recently, in a population-based screening program by Skaane et al.[22], the use of mammography plus DBT resulted in a significantly higher cancer detection rate and an increased detection of more invasive cancers compared to 2D only.

There is limited literature [12, 14, 23] that specifically addresses the clinical assessment of microcalcifications by DBT. FFDM appeared to be slightly more sensitive than DBT for the detection of calcification. However, diagnostic performance assessed using BI-RADS was not significantly different. The evaluation of microcalcifications in DBT is a topic under investigation with no conclusive results.

The first experience to work with DBT in the University Hospitals of Leuven was in 2009 with a Siemens Mammomat Inspiration system (Siemens, Erlangen, Germany). Approval of the ethics committee was obtained for DBT imaging. Informed consent was not required: the decision for DBT was left to the radiologist's clinical decision. The patients with very dense breasts and suspicious findings were referred to DBT after revision in 2D mammography. In 2010, it was noticed that some microcalcification clusters appeared differently in DBT compared to 2D which triggered the study as explained in chapter 7. In April 2010, a Hologic Selenia Dimensions DBT system (Hologic Inc. Bedford, MA, USA) was installed in the hospital. Currently, DBT is routinely used for difficult diagnostic cases.

1.2. Thesis objectives and methodological approach

Breast cancer remains a major health concern and a leading cause of cancer mortality among women [24, 25]. Mammography is an established method for breast cancer screening with a sensitivity

reported to be 83.5% [26]. This should be further improved. One major reason for missed cancers is the anatomical noise that can hide a tumour [1]. To overcome the shortcoming of overlapping tissues and therefore allow better lesion detection in breast screening and diagnosis, diagnostic radiology is embracing new 3D technologies such as DBT [1, 3, 8, 27]. As with all new technologies, additional investigations concerning the diagnostic or screening accuracy of the system, the technical quality and the effect of the different imaging parameters on the detection are necessary. Attempts to optimize and investigate DBT systems can be divided into three approaches: clinical trials or using phantoms or simulations. As presented earlier, both clinical trials or selective clinical studies are expensive in terms of time, money, patient safety and limited in the study set-up. Alternatively, the use of phantoms can overcome these limitations. Unfortunately, phantoms can be used to optimize only certain aspects since there exists no homologated phantom that can predict the clinical performance of DBT and certainly not when compared to 2D. That gave rise to dedicated simulation approaches [28] that can either be a full simulation approach [7, 28, 29], as in so-called virtual clinical trials, or a partial simulation approach in which observer studies are conducted on clinical images with simulated lesions [30–32]. After successful validation studies, simulation approaches might contribute to the optimization of different parameters, as they allow controlled conditions, which in turn might help to better separate benign from pathological findings. Therefore, the development of 3D models of breast lesions (microcalcifications and masses) with realistic appearance is crucially needed to support such studies.

We would like to be part of the present and also part of the future. For the present, we adopted the partial simulation approach to enable observer studies with acquired patient images and simulated lesions. For future applications, we focused on the development of 3D models of lesions to be included into software breast phantoms for future virtual clinical trials. The aim of this thesis was to develop a methodology to optimize the performance of breast tomosynthesis, our objectives were to:

1. Design and validate a partial simulation framework to simulate 3D lesions into DBT images.
2. Develop 3D models of microcalcification clusters and validate the realism of their appearance in 2D and DBT.
3. Develop 3D models of breast masses and validate the realism of their appearance in both 2D and DBT.

Accordingly, these objectives were worked out in the following chapters.

Chapter 2: We designed a partial simulation framework to simulate 3D objects into DBT images. The simulation started by identifying some input information including background material, insertion position, etc. The 3D object was then ray traced to calculate 2D projections/templates (i.e. x-ray transmission templates) taking into account the polychromatic x-ray spectrum. The 2D templates of the object were modified to include object sharpness then inserted at calculated positions in the projection images of the background while taking into account scatter corrections. These hybrid projections were then reconstructed. This framework was validated by comparing the contrast of simulated objects to their real ones. The objects used in the validation study were inserts found in phantoms. This work has been published in Radiation Protection Dosimetry (E. Shaheen et al. 2010 [33]).

Chapter 3: In this chapter, we proposed a novel method to develop 3D models of microcalcification clusters. The aim was to have a variety of shapes and to include benign and malignant clusters. Therefore, biopsy specimens of microcalcification clusters were collected and scanned by a micro-CT. The reconstructed images were then segmented to build the 3D models. A database of 3D models was created and the framework in chapter 2 was used to simulate them into 2D and DBT projection images. The realism of the appearance of these simulated clusters was validated in two separate observer studies that we report in this chapter. This database can be considered suitable for use in future observer performance studies. This work was published in Medical Physics (E. Shaheen et al. [34] 2011).

Chapter 4: In chapter 3, we developed 3D models of microcalcification clusters, but as stated in our objectives, it is not enough to conduct studies with only one type of lesion. Therefore, in this chapter, we report a new method of building 3D mass models of the breast with different morphological shapes. Two main categories of masses are described: non-spiculated and spiculated masses. To develop non-spiculated masses, contrast enhanced MRI breast lesions were collected, segmented, combined, triangularly meshed and scaled to different sizes. To create spiculated masses, the segmented masses were used as nuclei onto which spicules were grown. These mass models were projected into 2D and DBT images by means of the simulation framework in chapter 2. The realism of the appearance of these mass models was assessed by five radiologists in two separate validation studies. Once validated, this database is suitable for use in virtual clinical trials. This work was submitted to Medical Physics.

Chapter 5: This chapter introduced our first application to use the simulation framework presented in chapter 2. Microcalcification visualization in DBT is a hot topic under investigation with no conclusive results. Some concerns were noticed, in our images, suggesting that microcalcifications were less visible in DBT when compared to 2D. One major reason for the DBT system under investigation is that it is equipped with an x-ray tube that moves continuously during exposure causing blurring of small lesions due to the focal spot motion. This may hamper detectability and diagnosis of microcalcifications. In this study, we investigated the influence of position above the table on microcalcification contrast and signal difference to noise ratio (SdNR) in tomosynthesis images, by means of the simulation framework. This work was published in the Proceedings of SPIE Medical Imaging (E. Shaheen et al. [35] (2013))

Chapter 6: Adding to the concern mentioned earlier with the visibility of microcalcifications in DBT, there was a concern of misdiagnosing the microcalcifications in DBT. In this chapter, we present our first investigation of the morphological assessment of microcalcification clusters in 2D versus DBT. We focus on the morphology of calcifications because the morphological description of the shape of microcalcifications is a determining factor for recalling a screened woman or not. Four radiologists participated in the study and have described the shapes of microcalcifications in both simulated and real clusters in 2D images and DBT series that were read in separate blinded sessions. This work was published in the Proceedings of IWDM (E. Shaheen et al. [36] (2012)).

Chapter 7: Some concerns were stated during oral presentations of the previous chapter due to the use of simulated clusters for the morphological assessment and the lack of a known morphological descriptor. Therefore, in this chapter, we conducted a study to evaluate real microcalcification clusters in 2D when compared to DBT in terms of morphology and number of calcifications detected in each cluster. Six radiologists evaluated the clusters in 2D FFDM and DBT images. The two modalities were read in separate sessions to reduce bias. For each modality and for each cluster, the observers were asked to identify all the Le Gal types (morphology) present in the cluster and to count the number of visible microcalcifications. This work was submitted to British Journal of Radiology.

Chapter 1

Chapter 2

The simulation of 3D objects into breast tomosynthesis images

adapted from E. Shaheen et al., Radiat. Prot. Dosim. 139, 108–112 (2010)

Chapter 2

Abstract

Digital breast tomosynthesis (DBT) is a new 3D breast imaging modality that produces images of cross-sectional planes parallel to the detector plane from a limited number of x-ray projections over a limited angular range. Several technical and clinical parameters have not yet been completely optimized. Some of the open questions could be addressed experimentally, other parameter settings cannot be easily realized in practice and the associated optimization process requires therefore a theoretical approach. Rather than simulating the complete 3D imaging chain, we hypothesize that the simulation of small lesions into clinical (or test object) images can be of help in the optimization process. In this chapter, we describe how we simulated small 3D objects into real projection images. Subsequently these hybrid projection images are reconstructed using the routine clinical reconstruction tools. We report on the design and validation of this simulation framework via the comparison between simulated and real objects in reconstructed planes. The results confirm that there is no statistically significant difference between the simulated and the real objects. This suggests that other small mathematical or physiological objects could be simulated with the same approach.

2.1. Introduction

Digital breast tomosynthesis is being investigated as a new three dimensional (3D) imaging technique for early detection of breast cancer. As it provides volumetric information of the breast, it is considered a possible solution to overcome the main shortcoming in 2D conventional mammography: the overlap of anatomical structures on pathologies of interest [37]. The ability to reveal the third dimension is due to the acquisition of images of a stationary compressed breast at multiple angles during a short scan. These low dose projection images are then reconstructed into a series of thin high-resolution tomographic planes and can be displayed individually or in a cine mode.

The optimization of detector performance, acquisition geometry, acquisition parameters and reconstruction methods is an active topic of research and may be crucial for the acceptance of this new modality. The optimization of these factors should improve the clinical performance of the system and therefore should be done in terms of the detectability of lesions. Recently, simulation approaches are encouraged for the investigation of the new imaging modalities to overcome the limitations caused by the traditional clinical techniques as mentioned in chapter 1.

The aim of this work was to develop a 3D simulation framework to optimize and quantify the performance of DBT systems following the partial simulation approach. The process starts by

simulating small 3D objects into real projection images prior to reconstruction. These hybrid images are then reconstructed and can be used to study the detectability of the simulated lesions. It is the purpose of this chapter to validate the proposed simulation framework.

2.2. Materials and Methods

2.2.1. System specifications

The system used in this study was a Siemens Mammomat Inspiration with DBT functionality (Siemens, Erlangen, Germany), whose x-ray tube moves continuously over an angular range of ±25° in approximately 21 s and produces 25 projection images. The system is equipped with an *a*-Se detector of dimension 24 x 30 cm^2 with a pixel size of 85 μm. The source to detector (*a*-Se surface) distance (SID) is 65 cm with the center of rotation (COR) of the x-ray tube 4 cm above the table top (Fig. 2.1).

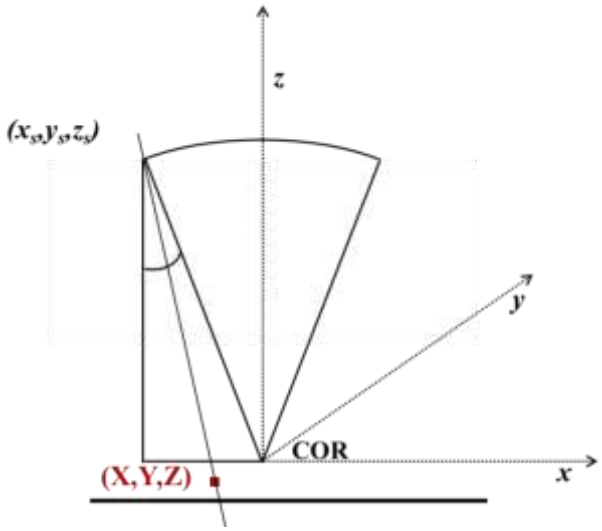


Figure 2.1 Geometry of the tomosynthesis system used

2.2.2. Simulation process

The simulation procedure, as detailed in Fig. 2.2, is divided into three main steps: *1. Object simulation*: the geometrical information is retrieved to calculate the attenuation of the primary x-ray beam for the theoretical object. Primary x-ray attenuation is calculated with ray tracing and spectrum simulation to create "2D templates" (Eq. 2.1). *2. Template modification:* the templates are modified to include the system resolution characteristics, noise and scatter. *3. Insertion and reconstruction:* the

templates are inserted into their pre-defined locations in the projection images and the hybrid images are reconstructed using the routine reconstruction software. The next sub-section describes this in more details. All the simulation steps included in the study were implemented using MATLAB (version 7.8, MathWorks, Natick, MA).

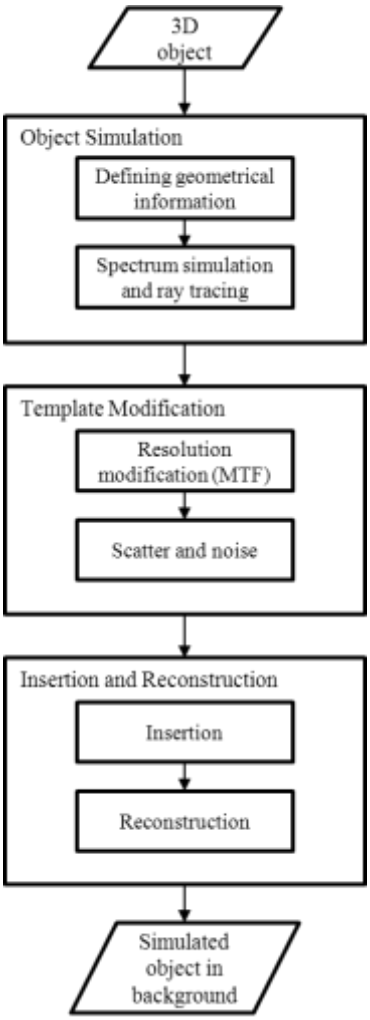


Figure 2.2 Simulation process flow chart

2.2.2.1. Object simulation

The 2D templates are obtained through spectrum simulation and ray tracing. The polychromatic spectrum was simulated using the narrow beam spectral model of Boone *et al.* [38] and the ideal templates were subsequently obtained by integrating the energy specific attenuation of the different materials (i.e. the support plate, the detector cover, the phantom/patient and the object) over the complete spectrum as described in Eq. 2.1. The linear attenuation coefficients of the different

materials were retrieved from the attenuation data of the National Institute of Standards and Technology (NIST) [39] and XMUDAT software [40]. The thickness of the object was calculated by ray tracing the object model using the Siddon algorithm [41]. To do so, the geometrical information of the system, the 3D insertion position (X, Y, Z) (i.e. the theoretical position of the object in the background phantom/patient relative to the geometry of the system) and the projection angles were required (Fig. 2.1).

$$temp_{ideal}(x,y) = \frac{\sum_E E \in (E) N(E) e^{-\sum_i \mu_i(E) t_i - \mu_{back}(E) t_{back} - \mu_{obj}(E) t_{obj} + \mu_{back}(E) t_{obj}}}{\sum_E E \in (E) N(E) e^{-\sum_i \mu_i(E) t_i - \mu_{back}(E) t_{back}}} \quad (2.1)$$

where E is the incident photon energy, ∈ (E) is the detector energy absorption efficiency, N(E) is the relative number of x-ray photons at energy E calculated from the spectral model of Boone et al., μ_i is for the linear attenuation coefficients of the different traversed materials (the support plate, the detector cover…) and the corresponding thicknesses t_i, μ_{back} is the linear attenuation coefficient of the background material (patient/phantom), t_{back} is the thickness of the background, μ_{obj} is the linear attenuation coefficient of the simulated object, t_{obj} is the thickness of the object traversed by the ray at this position.

2.2.2.2. Template modification

The 2D templates were modified for system resolution characteristics by means of the modulation transfer function (MTF). The pre-sampled MTF was measured based on the edge method [42] using a steel edge of 0.8 mm thickness placed on the breast support platform as described by Marshall [43]. The MTF curves were measured in the front-back and tube-travel directions from the central projection image (~0°) of the edge for the tomosynthesis mode. Therefore, the MTF in the tube-travel direction included information on blurring from the tube motion in the left-right direction across the detector. MTF in the front-back direction across the detector is unaffected by the tube motion. The 2D MTF function for every projection angle was generated by linear interpolation between the MTF in the tube travel direction and the curve for the front-back direction. This 2D MTF (MTF_{2D}) function was multiplied in the frequency domain with the Fourier transformed (*FT*) template as described in Eq. 2.2 and the subsequent inverse Fourier transform (*IFT*) resulted into the modified template.

$$temp_{modified}(x,y) = IFT(FT(temp_{modified}(x,y)) \cdot MTF_{2D}) \quad (2.2)$$

The subtle changes in total attenuation in the region of insertion with or without a small object are unlikely to affect the noise statistics of the region surrounding the object. Therefore, the noise was assumed to be inherited from the background region.

2.2.2.3. Insertion and Reconstruction

The templates are intended for insertion into the raw data of the real projection images of a phantom or patient. Due to the fact that the templates describe the attenuation of the primary x-rays only (scattered radiation is not altered), they need to be inserted into the primary part of the projection images. The DBT system used in this study does not have an anti-scatter grid, and this means that scattered radiation must be accounted for in the simulation framework. In a first approach, the scatter-to-primary ratio (SPR) for every projection angle was obtained based on the values reported by Sechopoulos et al. [44], with the equivalent compressed breast thickness reported in the DICOM header used as input. The scatter is known to be uniform in the central part of the breast [44]; all the objects reported here have been simulated in the central region of the phantoms. Other SPR values should be used in more peripheral parts. This SPR value is utilized to calculate the scatter fraction (SF) as shown in Eq. 2.3.

$$SF = \frac{SPR}{SPR+1} \tag{2.3}$$

A considerably large region of interest (ROI) around the object is defined in the raw data. Using the SF, the primary component is estimated by subtracting the scatter offset ($SF.\bar{I}$) from the original pixel values of the ROI ($I(x,y)$). The insertion is done by simple multiplication of every modified template ($temp_{modified}(x,y)$) by the corresponding estimated primary component for the ROI. The scatter offset is then added back. The insertion procedure uses the following equation (Eq. 2.4):

$$I'(x,y) = \left((I(x,y) - SF.\bar{I}).temp_{modified}(x,y)\right) + SF.\bar{I} \tag{2.4}$$

These hybrid tomosynthesis projections were subsequently reconstructed using the Siemens software (TomoEngine, Siemens, Erlangen, Germany) that is based on the filtered back-projection algorithm (FBP) [45]. The reconstructed voxel size was 0.085x0.085x1.0 mm^3.

2.2.3. Validation study

The validation of the simulation approach required the availability of test objects with well-defined geometrical inserts in terms of shape, size and material. The inserts of three phantoms were used. They are described in Table 2.1 and Fig. 2.3.

The gold disks from a CDMAM 3.4 phantom (Artinis, The Netherlands) were used as objects (inserts) for testing. The diameters and thicknesses information was used to simulate the objects into the projection images of a CDMAM phantom of identical make but without gold disks. Identical exposure settings had been applied to both phantoms (CDMAM with and without disks). These images had the same characteristics apart from the presence of the gold disks. The CDMAM acquisitions were performed in combination with different thicknesses of Polymethyl methacrylate (PMMA) slabs, typically 2, 4 and 6 cm, providing three different datasets. Within these datasets, a total of 120 disks were simulated in the centre of the (empty) squares.

Cylinders in a CDRAD 2.0 phantom (Artinis, The Netherlands) were used as inserts. These cylinders are of air (holes in the shape of cylinders) and were simulated into the projection images of the phantom in the same cell next to the real cylinder. The phantom was placed on top of 2cm of PMMA and another 1cm of PMMA was placed on top of it to make a total of 4cm PMMA. The dataset contained 25 elements and simulated and real cylinders with the same nominative size were compared.

Four spheres of breast simulating material (CIRS, Norfolk VA, USA) with different glandularity (density) percentages (0%, 30% , 50%, 100%) in an oil phantom were simulated and compared to real acquisitions.

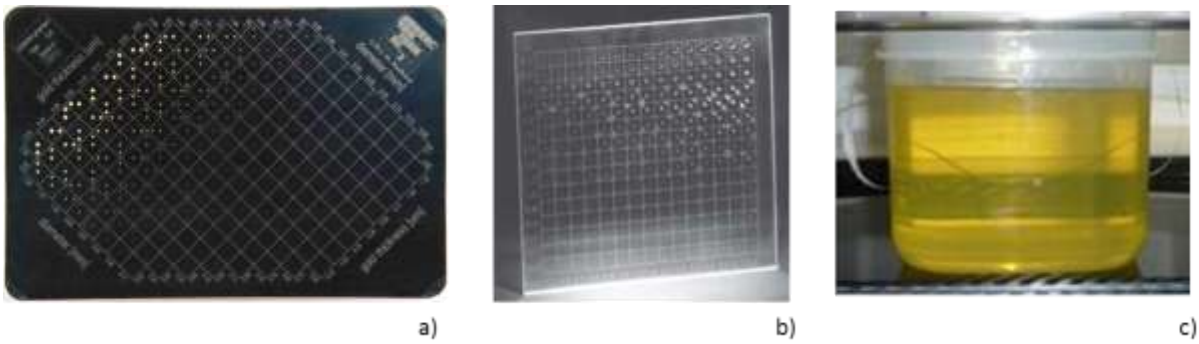


Figure 2.3 The phantoms used in the validation study.
a) CDMAM 3.4 phantom, b) CDRAD 2.0 phantom, c) CIRS sphere in oil

Table 2.1 The objects used in the validation study including the phantoms, the type of object, the number of objects and description.

Dataset	Phantom	Objects (inserts)	Description	No. of objects
1	CDMAM 3.4	Gold disks	cylinders with known shape, size (diameter, thickness). Material: gold.	120
2	CDRAD 2.0	Cylinders of air	cylinders with known shape, size (diameter, thickness). Material: air.	25
3	CIRS spheres in oil	Spheres of breast simulating material	Breast simulating materials of 4 glandularity percentages (0%, 30%, 50%, 100%) of 5 mm diameter	4

The contrast of the object relative to its background (Eq. 2.5) was our figure of merit to compare the real and simulated objects. The contrast was measured for each simulated object and its corresponding real object. A region of interest (ROI) was defined as a circle with slightly smaller diameter than the diameter of the object and centered in the center of the object to measure the signal inside of the object. Another ROI was defined as a circle with a larger diameter taken within the same cell in a corner, as shown in Fig. 2.4 to measure the background signal. The diameter of the background ROI was large enough to give a good estimate of the background signal, and in the corner to guarantee no reconstruction artifacts were included. Then these series of contrast values in all the simulated and corresponding real objects were compared with a two sided t-test for each dataset. P-value > 0.05 indicates no statistically significant differences between simulated and real objects.

$$C = \frac{|m_{signal} - m_{back}|}{m_{back}} \tag{2.5}$$

where m_{signal} is the mean pixel value in the object ROI, m_{back} is the mean pixel value in the background ROI.

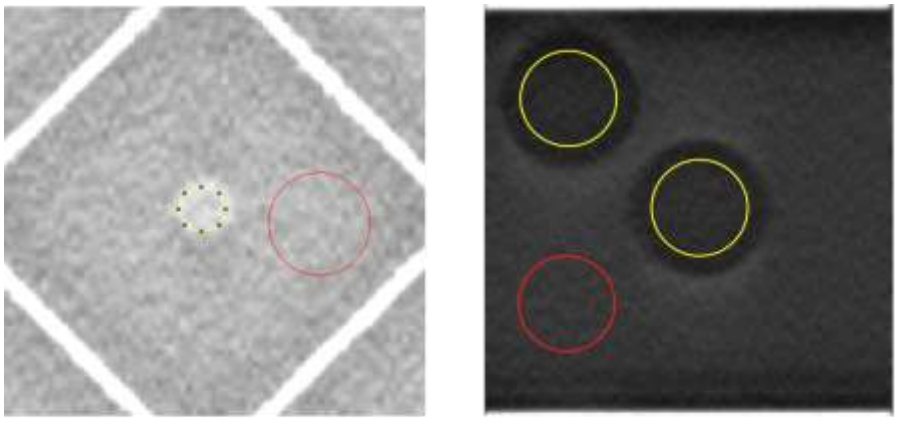


Fig. 2.4 Contrast measurements for CDMAM disks (a) and CDRAD cylinders (b). The yellow circle represents the signal ROI and the red circle represents the background ROI

2.3. Results

The contrast measurements were applied to both the simulated and the corresponding real objects as shown in Fig. 2.5 for all datasets. The compared contrast values fit linearly: for the 120 CDMAM disks, the p-value was 0.32, for the CDRAD dataset, the p-value was 0.51. No significant difference between the simulated and the real objects in terms of inherited contrast was found. For the spheres of the breast simulating materials, high correlation was found, but due to the small dataset, no p-value was retrieved.

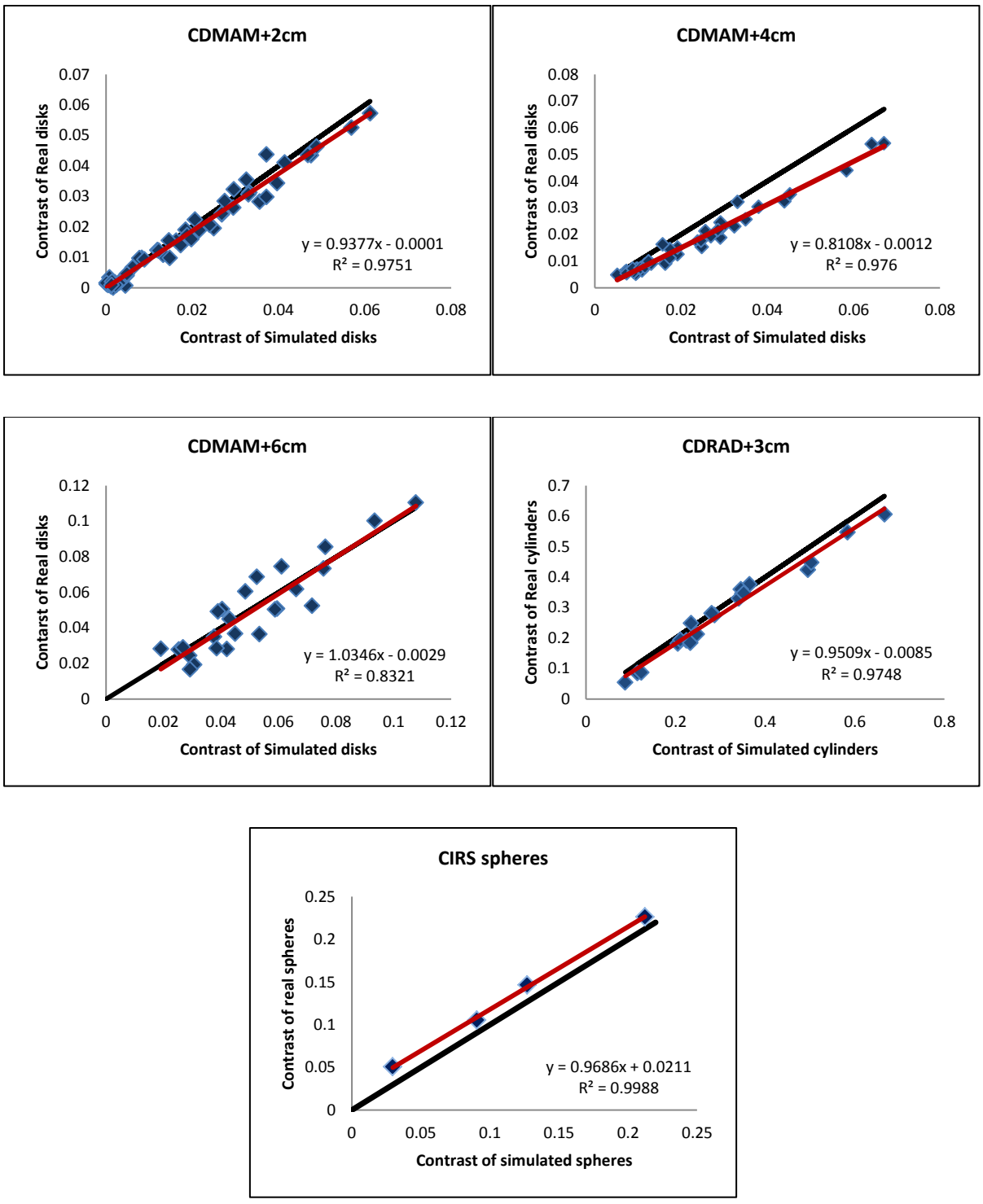


Fig. 2.5 Contrast of simulated versus real objects for the three datasets of CDRAD, CDRAD and CIRS spheres.

2.4. Conclusion

The developed simulation framework was validated by means of comparing the contrast of simulated inserts of phantoms to the contrast of the acquired inserts in the reconstructed images. As shown in the results, there was no statistically significant difference in terms of contrast between simulated and real objects. This framework can assist phantom designs by testing different materials, objects and positions in the background phantom. Further development aims to extend this work to simulate 3D breast lesions in DBT images and perform comparative observer performance experiments such as the comparison of different reconstruction algorithms in terms of lesion detection.

ACKNOWLEDGMENT

To Thomas Mertelmeier (Siemens, Erlangen, Germany) for reconstruction software and discussions about the simulation procedure; to Roeland Vanderburght to make the 'CDMAM without gold disks' available; to Alain Noël (Nancy, France), for sharing his experience with simulating 2D CDMAM images. This work was financially supported by the EC-EURATOM 7 Framework Program and forms part of the project on "Dedicated CT of the Female Breast: Feasibility, Optimisation and Comparison to competing imaging modalities", contract no. FP/213153.

Chapter 2

Chapter 3

The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis

adapted from E. Shaheen et al., Med. Phys. 38(12), 6659-6671 (2011).

Chapter 3

Abstract

Purpose: This work proposes a new method of building 3D models of microcalcification clusters and describes the validation of their realistic appearance when simulated into 2D digital mammograms and into breast tomosynthesis images.

Methods: A micro-CT unit was used to scan 23 breast biopsy specimens of microcalcification clusters with malignant and benign characteristics and their 3D reconstructed datasets were segmented to obtain 3D models of microcalcification clusters. These models were then adjusted for the x-ray spectrum used and for the system resolution and simulated into 2D projection images to obtain mammograms after image processing and into tomographic sequences of projection images, which were then reconstructed to form 3D tomosynthesis datasets. Six radiologists were asked to distinguish between 40 real and 40 simulated clusters of microcalcifications in two separate studies on 2D mammography and tomosynthesis datasets. Receiver operating characteristic (ROC) analysis was used to test the ability of each observer to distinguish between simulated and real microcalcification clusters. The kappa statistic was applied to assess how often the individual simulated and real microcalcification clusters had received similar scores ("agreement") on their realistic appearance in both modalities. This analysis was performed for all readers and for the real and the simulated group of microcalcification clusters separately. "Poor" agreement would reflect radiologists' confusion between simulated and real clusters, i.e. lesions not systematically evaluated in both modalities as either simulated or real, and would therefore be interpreted as a success of the present models.

Results: The area under the ROC curve, averaged over the observers, was 0.55 (95% confidence interval [0.44, 0.66]) for the 2D study, and 0.46 (95% confidence interval [0.29, 0.64]) for the tomosynthesis study, indicating no statistically significant difference between real and simulated lesions ($p>0.05$). Agreement between allocated lesion scores for 2D mammography and those for the tomosynthesis series was "poor".

Conclusions: The realistic appearance of the 3D models of microcalcification clusters, whether malignant or benign clusters, was confirmed for 2D digital mammography images and the breast tomosynthesis datasets; this database of clusters is suitable for use in future observer performance studies related to the detectability of microcalcification clusters. Such studies include comparing 2D digital mammography to breast tomosynthesis and comparing different reconstruction algorithms.

3.1. Introduction

The ability of digital breast tomosynthesis (DBT) to enable detection and categorization of microcalcifications remains uncertain today [14]. While some studies [18, 23] have compared the equivalence or superiority of DBT to full-field digital mammography (FFDM) in terms of cancer detection, further research is required to fully characterize differences in detection performance between modalities. Microcalcifications are of fundamental importance in this regard. As the morphology of microcalcifications is a determining sign of malignancy [46, 47], it is important to study several aspects of these lesions in much detail.

Some studies have addressed the topic of simulating microcalcifications in FFDM [48–50] but experience in DBT is still limited [7]. If studies with simulated microcalcifications are to provide a true representation of observer performance for a given modality, it is important that these lesions are indistinguishable from real microcalcifications for observers. This requires validation of the methodology used to simulate these lesions and confirmation of their realistic appearance [48, 51]. A crucial aspect of a correct simulation procedure is the preservation of microcalcification morphology to keep the realistic appearance of these simulated lesions when used in radiological observer studies and to avoid biasing the judgment of the readers participating in the studies.

The purpose of this study was the development of accurate three-dimensional (3D) models of microcalcification clusters with a variety of cluster types. A further aim was to validate the realistic appearance of these models when simulated and inserted into patient images acquired using a two-dimensional (2D) FFDM system and a breast tomosynthesis unit. To this end, an observer study in which radiologists were asked to distinguish between real and simulated clusters was conducted separately for each modality.

3.2. Materials and Methods

The simulation of microcalcification clusters started with the collection of breast biopsy specimens that were scanned by a micro-CT unit and then segmented. An initial set of 3D models of microcalcification clusters was formed and a number of clusters were used later in two validation studies, as shown in Fig. 3.1. The simulation framework [33], described in chapter 2, was applied to these models to simulate them into projection images. Their realistic appearance was then validated for 2D FFDM and DBT.

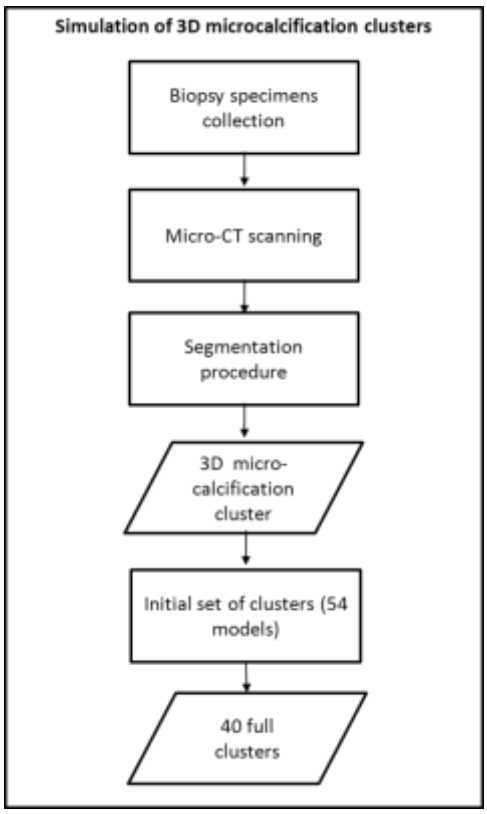


Fig. 3.1 A flow chart that explains the main steps to acquire 3D models of microcalcification clusters

3.2.1. Simulation of 3D models of microcalcification

3.2.1.1 3D model of microcalcification cluster

The 3D models of microcalcification clusters were built from biopsy specimens; a photograph of a biopsy specimen is shown in Fig. 3.2 (a). The specimens were obtained with a stereotactic guided vacuum large core needle (10 gauge) by means of a digital biopsy table (Mammotest Plus/S, Fisher Imaging, Norderstedt, Germany). For every patient case, the sample with the microcalcifications was preserved in a tube with Formaldehyde (CH_2O) and was imaged using a cone beam micro-CT scanner (SkyScan 1172, SkyScan, Aartselaar, Belgium) that provides projection and 3D reconstructed images.

The micro-CT exposure settings were adjusted for each sample. Tube voltage was varied from 60 to 80 kVp and with the tube current time product (µAs) adjusted accordingly following a pre-defined set exposure table. For example, when the tube voltage was 60 kVp, 167 µAs was used. The scanning pixel size ranged from 17 to 30 µm depending on the shape (regular, irregular), size of the microcalcifications (very small, small, normal) and on the size of the whole cluster in the sample. In

some cases, reconstruction artifacts were present with the default 20 µm acquisition set, and then other pixel sizes or other tube voltages were chosen by trial and error to overcome this problem. Isotropic voxels were reconstructed. The microcalcification cluster models were constructed by segmenting the micro-CT 3D images.

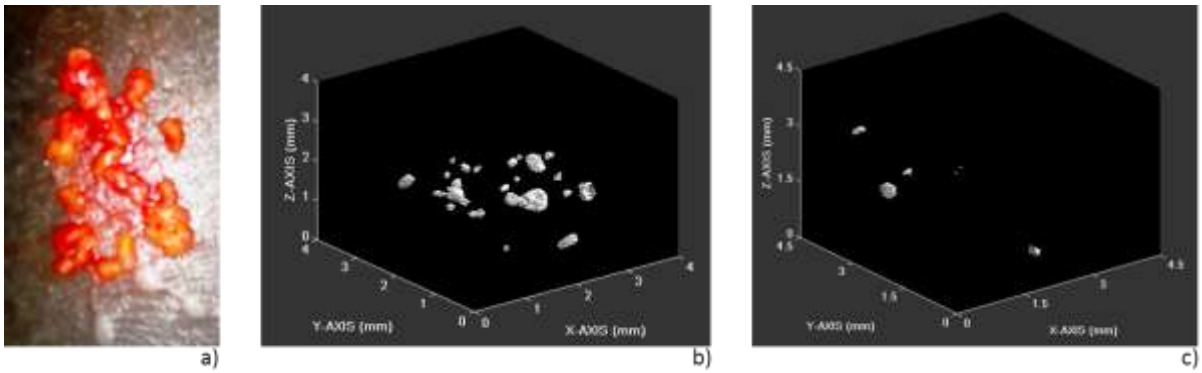
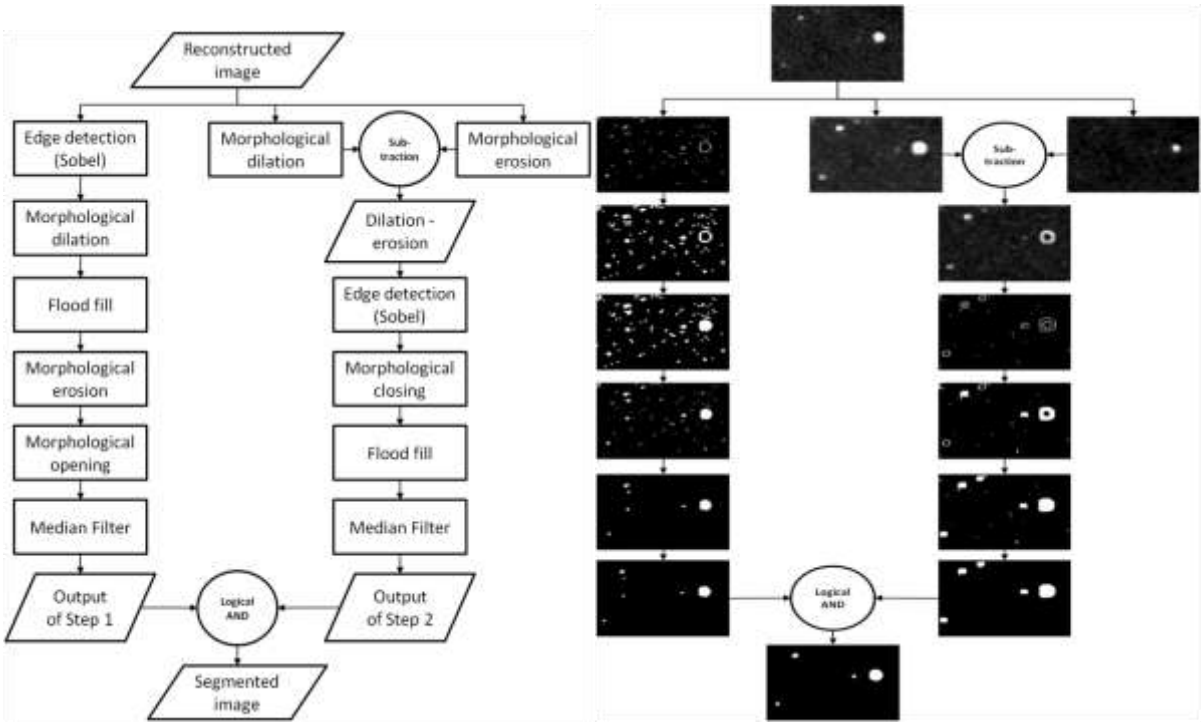


Fig. 3.2 a) A photograph of a fresh biopsy specimen just after extraction from a patient, the specimen contained microcalcification clusters (not visible in this image). b) An example of a 3D model of microcalcification cluster from the database built during this study and considered as full cluster (voxel size was 0.02 mm). c) An example of a 3D model of microcalcification cluster from the database and considered as sub-cluster (voxel size was 0.03 mm).

A segmentation procedure was applied to every reconstructed image of a biopsy specimen as shown in Fig. 3.3. This procedure was divided into two separate algorithms and their outputs were combined through a logical AND function to produce the final output. The overall procedure is described in Fig. 3.3 (a) with the detailed steps within each algorithm while Fig. 3.3 (b) shows the identical steps but illustrated with figures. The first algorithm consisted of edge detection using the Sobel operator followed by a set of morphological operations: dilation, erosion and opening, where the structuring element for the erosion and dilation was of a disk shape with radius = 1 voxel and the structuring element of the opening had a linear shape of size 1 x 3 voxels that was applied once vertically then horizontally. Other operations were also included such as flood fill and a median filter of size 3 x 3 voxels. The second algorithm included the application of the morphological gradient transform filter (i.e. subtracting the output of a morphological erosion operation from the output of a dilation operation applied on the original image), edge detection using the Sobel operator and morphological closing. All morphological operations had a disk shape as structuring element with a radius of 3 voxels, followed by flood fill and median filter of size 3 x 3 voxels. The output of each algorithm and the final output are illustrated in Fig. 3.4 where the original image (the micro-CT reconstructed image) is shown with the borders of the segmentation overlaid on the image in red to explain the effect of the combination step through logical AND. The first algorithm provided better segmentation

of the boundaries of the microcalcifications but suffered from background noise, as shown in Fig. 3.4 (a). While in the second algorithm, the borders were segmented larger than the true borders but had fewer false positives, i.e. noise with the appearance of microcalcifications (Fig. 3.4 (b)). The binary outputs of the first and second algorithm were combined through logical AND and the final result was an image with accurately segmented boundaries and fewer false positives as shown in Fig. 3.4 (c). The output of the segmentation procedure was a 3D binary model where the background voxels were set to zero and the voxels within the object had a value of one (Fig. 3.3 (b)).



a) b)

Fig. 3.3 a) A flow chart explaining the steps involved in the segmentation procedure of the micro-CT reconstruction images to form a 3D model. b) All the steps in a) are illustrated by images as output of each step.

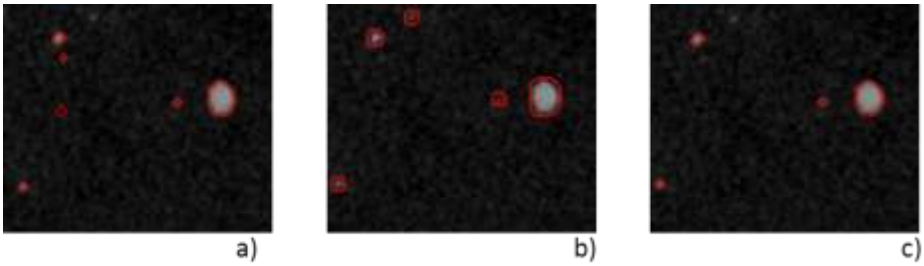


Fig. 3.4 The different outputs of the segmentation procedure applied on each reconstructed micro-CT image shown for visual check with the original input in gray scale and the borders of the output represented by the red contour. a) The output of algorithm 1 in the segmentation procedure. b) The output of algorithm 2 in the segmentation procedure. c) The final output of the segmentation procedure after the combination of algorithms 1 and 2 through logical AND.

3.2.1.2. Initial sets of microcalcification clusters

An initial number of 23 biopsy specimens was available. Based on the pathology reports, 13 were found to be malignant and 10 were benign. Since a cluster of microcalcifications is usually composed of different types of microcalcifications where some could be benign and some malignant, the decision of benign is taken when all microcalcifications in the cluster are benign. On the other hand, a cluster is considered malignant with the presence of at least one malignant microcalcification. In order to extend this initial set of microcalcification clusters, every model was rotated in different directions producing three copies of each model. The projection image was calculated for each copy. When the 2D distributions of these microcalcifications within the copy were sufficiently different from the distributions in the original model (judged by simple visual inspection by the scientist involved with the study and an experienced radiologist), this copy was considered a new model. This procedure increased the number of models from 23 to 54.

Based on the experience of the radiologist involved with the stereotactic biopsy procedures in our hospital, it was decided to divide the microcalcification clusters into sub-clusters and full clusters. A sub-cluster was defined as a group of maximum six microcalcifications that would not be considered representative for a stand-alone cluster. Full clusters consisted of seven or more microcalcifications. They can be used in stand-alone mode and applied immediately for simulation purpose. The 54 microcalcification clusters were classified into 34 full clusters and 20 sub-clusters by consensus reading of radiologist and scientist in charge of the current project. The main reason for the existence of sub-clusters was due to the vacuum assisted biopsy procedure where a large cluster was divided into several groups of microcalcifications in different specimens. Multiple scans were then required due to the limited micro-CT field of view, resulting in sub-clusters rather than one full cluster. When a sub-cluster was included in the validation study, it was combined with another sub-cluster of the same biopsy procedure to give a subjectively realistic appearance. Examples of 3D models of microcalcification clusters are shown in Fig. 3.2.

3.2.2. Simulation framework

The simulation framework, that was presented in chapter 2, was previously designed and validated to simulate small 3D objects into the projection images of a breast tomosynthesis system [33]. The same framework was used to simulate objects in 2D FFDM by confining the calculations to the 0° angle only and using the proper technical characteristics. Differences between the simulation methods for the two modalities and the modifications applied to the framework in this study compared to the one

developed in chapter 2 are highlighted in the next sections. All the simulation steps included in the study were implemented using MATLAB (version 7.8, MathWorks, Natick, MA). The system specifications in section 2.2.1 are the same used in this study.

For the object simulation part, the input objects were the 3D microcalcification clusters and calcium oxalate (CaC_2O_4) was used as substitute material for calcifications [52]. For the template modification part, the MTF curves were measured in the front-back and tube-travel directions from the central projection image (~0°) for the tomosynthesis mode. In the case of 2D digital mammography, the MTF was measured in the same way using the 2D acquisition mode. In order to obtain the MTF for the oblique incidence of the x-rays to the detector surface in case of the tomosynthesis mode, the model described by Que and Rowlands [53] was applied to the MTF curve in the tube travel direction for every projection angle. The 2D MTF function was multiplied in the frequency domain with the Fourier transformed template as described in section 2.2.2.2.

Microcalcifications are quite highly attenuating; therefore, the photon statistics in a region within the calcification may be changed. Microcalcifications have, however, small physical extent in the images and are unlikely to affect the noise statistics of the region surrounding the microcalcification. Therefore, noise was not added to the images in regions adjacent to the calcification; the noise was inherited from the background region.

The SPR values were taken from Boone's work [54]. The spatial distribution of scattered radiation is known to be fairly uniform within the central part of the breast [44] and hence the insertion positions were limited within 80% of the central portion of the breast, away from skin edge borders. To ensure a realistic simulation of microcalcifications within the breast area, the locations of insertion were further chosen based on the clinical knowledge of the distribution of breast carcinoma [55]. The main parameters affecting the calculation of the SPR values are the compressed breast thickness and the oblique entry [44]. The compressed breast thickness was exported from the DICOM header and an increased breast thickness is calculated based on the corresponding oblique projection angle for the central projection ray. For example, a 4 cm thick breast with the geometry of the system used in the study and an oblique projection angle of 20° was assumed to have a new breast thickness of 4.25 cm. This "central oblique" breast thickness was used to estimate the SPR value for the corresponding projection angle [56]. The values of the grid transmission factors for primary (0.75) and scatter (0.14) for 2D digital mammography were obtained from measurements with lead beam stops and used to estimate the correct SPR value after transmission through the grid [57]. The templates were multiplied with the estimated primary part of the real tomographic input images as explained in Eq. 2.4. For

insertion in the 2D case, the SPR value after transmission through the grid is used and the template was multiplied with the estimated primary part of the raw projection input image as in Eq. 2.4.

These hybrid tomosynthesis projections were subsequently reconstructed using the Siemens software (TomoEngine, Siemens, Erlangen, Germany) that is based on the Filtered Back-projection algorithm (FBP) [45]. The reconstructed voxel size was 0.085 mm x 0.085 mm x 1 mm. For the 2D case, the default image processing algorithm for 2D digital mammography (OpView2, Siemens, Erlangen) was applied to the raw image to obtain a processed image. These reconstructed and processed images were used for the evaluation studies.

3.2.3. Validation of the simulation

3.2.3.1. Image datasets

Image data from a total of 67 patients were used in the validation studies (2D FFDM and tomosynthesis). Each patient was imaged in 2D and tomosynthesis using the same mammography system. Only the mediolateral (MLO) views of tomosynthesis and 2D mammography were used. All images were available in raw format. Thirty cases out of the 67 cases contained a total of 40 microcalcification clusters; these 40 clusters were used to form the group of real clusters. Forty microcalcification clusters were simulated into the images of the 37 patient cases without real microcalcification clusters, however some of these cases contained masses. A patient case with real clusters was never used as background for simulation. All clusters that had been simulated into tomosynthesis projections were then simulated into 2D mammograms of the same patient in the same location allowing a paired analysis. The distribution of the clusters and the patients used in the validation studies are described in Table 3.1.

The 40 simulated microcalcification clusters were selected from the set of 3D models mentioned in section 3.2.1.2. A total of 44 models were used to construct the 40 full cluster models participating in the validation studies where 34 models were chosen from the group of full clusters (stand-alone) and 6 other full clusters were built from sub-clusters as explained earlier. Details of the distribution and malignancy/benign description of these models are shown in Table 3.2.

Table 3.1. The description of the patients and clusters used in the validation studies and the training studies in terms of real and simulated population along with the total number of cases and clusters used in the overall study. The patients in the "group with real clusters" are patient cases with real microcalcification clusters while the patients within the "group with simulated clusters" are real patient cases where their background were used for simulating microcalcification clusters.

			Group of real clusters	**Group of simulated clusters**	**Total**
Validation studies		Patients	30	37	67
		Clusters	40	40	80
Training sessions	1. Learning to work with tomosynthesis	Patients	10	-	10
		Clusters	>10	-	>10
	2. Training on the software	Patients	5	5	10
		Clusters	5	5	10
Total (validation + training)		Patients	45	42	87
		Clusters	>55	45	>100

Table 3.2. The description of the models used in the validation studies in terms of quantities, distribution and malignancy/benign description

	Type of cluster	**Number of clusters**	**Number of malignant**	**Number of benign**
Models from initial set used in the validation studies	Full cluster	34	29	5
	Sub-cluster	10	5	5
	All clusters	44	34	10
Models participated in the validation studies	Full cluster (stand-alone)	34	29	5
	Full cluster formed from sub-clusters	6	4	2
	All full clusters	40	33	7

3.2.3.2. Observer study

Two observer studies were conducted independently, one for 2D digital mammography, the other for breast tomosynthesis. The quality and clinical utility of simulated microcalcification clusters was evaluated in terms of the ability of the radiologists to distinguish between simulated and real clusters. Receiver operating characteristic (ROC) analysis [58] was performed in order to assess the readers' ability to tell simulated and real clusters apart. While the standard use of the ROC methodology aims at differentiating between two states (e.g. normal and abnormal), in the current study the ROC method was used to assess the analogy (and not the difference) between simulated and real clusters. Simulated clusters of microcalcifications were considered as 'abnormal cases' in standard ROC analysis while real clusters were considered as 'normal cases'. Radiologists were asked to rate their confidence that a cluster was real on a 5-point scale (the scoring was from 1 to 5): 1. cluster is definitely real, 2. cluster is probably real, 3. cluster is possibly real, 4. cluster is probably simulated, 5. cluster is definitely simulated. These points are also called threshold points that are used to define the ROC curves. Under the null hypothesis (NH) that the real and simulated lesions cannot be distinguished, the area under the ROC curve = 0.5. Similar studies have followed this approach [51, 59].

Six radiologists participated in both studies with five having more than ten years of experience in screening and diagnostic mammography and the sixth radiologist a final year student under training. Three out of the five experienced radiologists had worked with breast tomosynthesis for about two years. A training phase, divided into two parts, was implemented before starting the studies: 1. learning to work with tomosynthesis; 2. training on the software.

1. Learning to work with tomosynthesis: ten patient cases, having at least one real microcalcification cluster per patient, were assigned to this training session. The images were read by each radiologist from the PACS system (see Table 3.1). Each patient case included MLO views of both 2D digital mammograms and tomosynthesis series and occasionally both craniocaudal (CC) and MLO views were available. The goal of this session was to train the radiologist to read the tomosynthesis series, to be familiar with the appearance of single microcalcifications and clusters of microcalcifications in tomosynthesis reconstructed images, to compare these microcalcifications to 2D digital mammography and to gain confidence with this new modality. The radiologist was free to select/change the image magnification and the window width and level. They were also free to discuss these cases with the other radiologists.

2. Training on the software: a user friendly software package SARA [60] that was developed earlier in the group was used as a visualization tool in the observer studies. This training session was divided into two sessions, one for 2D and the other for the tomosynthesis study. Ten different patient cases were included in these sessions; five real and five simulated microcalcification clusters were randomly distributed within the session (see Table 3.1). The aim of these sessions was to train the radiologists in the use of the software and in answering the question accompanying the studies. These sessions also served as a small pilot study for the simulation work. Feedback from the radiologists was used to improve the main study, principally in terms of better insertion positions of the microcalcifications clusters.

The validation studies were performed with a high resolution 5 megapixel monitor (Barco MDNG5121CB) in a routine mammography reading room. For the 2D study, the 80 real and simulated clusters were evenly divided into two reading sessions with the order of the clusters randomly distributed within the sessions. The images were presented to the radiologists one by one having the whole breast (only the MLO view) shown with a rectangle that could be toggled on and off around the cluster in question. The radiologists were free to zoom in/out, move within the mammogram and change the window level with no time limitations.

For the tomosynthesis study, the 80 clusters were evenly divided into four reading sessions with the order of the clusters also randomly distributed within the sessions. For each case, 15 reconstructed planes around the central "in-focus" plane of the cluster in question (+ 7 planes before and 7 planes after) were presented (1 mm plane spacing). The radiologists were asked to evaluate the whole cluster with the freedom to scroll within the 15 planes, zooming and changing window level and width. This mimicked the clinical situation in which the radiologist can also investigate the appearance of the cluster in different planes, where out of plane artifacts might affect the reader's judgment. The choice of 15 planes was a compromise between showing the complete set of out of plane artifacts and the speed of the readings. An example of out of plane artifacts is shown in Fig. 3.5; it can be seen that this relatively large microcalcification does not spread beyond 7 planes.

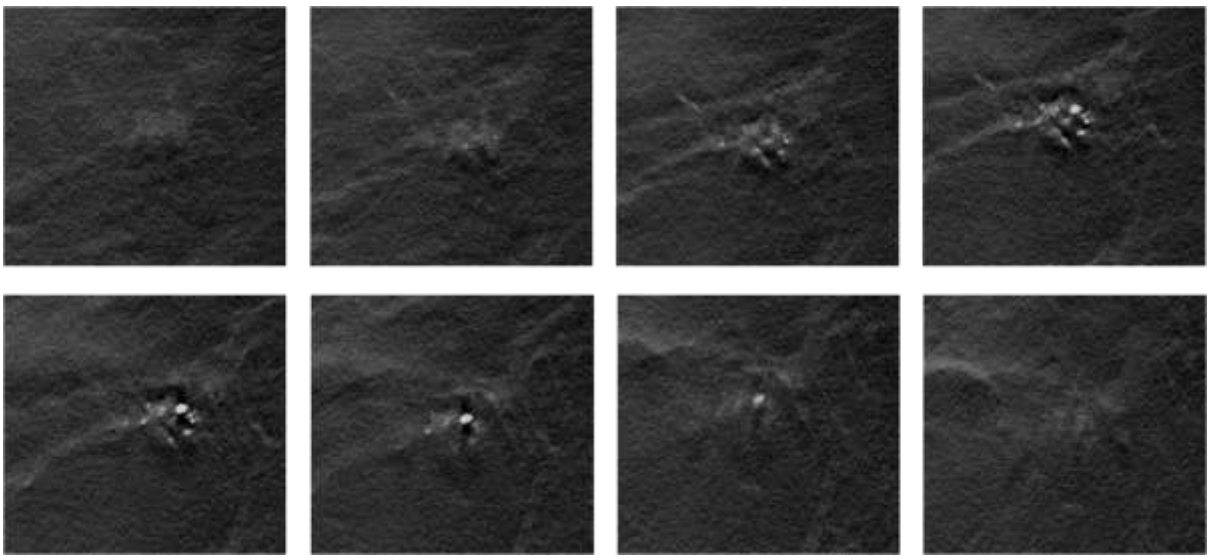


Fig. 3.5. Illustration of the out of plane artifacts of a real microcalcification cluster. These are 8 out of the 15 planes shown to the radiologists starting from the top left as 1, 3, 5…15.

3.2.3.3 Statistical analysis

DBM MRMC version 2.2 (Medical Image Perception Laboratory at University of Iowa) was used to analyze the results. The areas under the ROC curves (AUC) were computed using the trapezoidal method along with the 95% confidence interval (CI) for each reader and for each modality. An area under the ROC curve, of a particular radiologist, close to the chance value of 0.5 would indicate that the radiologist could not distinguish between real and simulated lesions. An average AUC, defined as the average of all readers' AUCs, along with the corresponding 95% CI was also calculated. If the 95% CI of the AUC includes the value 0.5, the null hypothesis of simulated and real lesions being undistinguishable cannot be rejected.

To measure the agreement between scores to the microcalcifications in 2D FFDM and DBT images, the commonly used Cohen's kappa statistic was used. The aim of this analysis was to check whether there was agreement between the 2D FFDM and tomosynthesis scores as a whole. The analysis included also an agreement test between the two modalities within the group of simulated clusters and within the group of real clusters. For each reader, the scores for both simulated and real clusters (separately) were used to calculate the weighted kappa statistic (κ) and its 95% CI (MEDCALC version 11.5.1.0, MedCalc Software, Ostend, Belgium). By convention, $\kappa < 0.2$ is interpreted as "poor" agreement, $0.21 < \kappa < 0.40$ as "fair" agreement, $0.41 < \kappa < 0.60$ as "moderate" agreement, $0.61 < \kappa < 0.80$ as "good" agreement, $0.81 < \kappa < 1.0$ as "very good" agreement.

3.3. Results

3.3.1. Results of the 2D FFDM study

Results of the ROC analysis per observer are shown in Table 3.3 and Fig. 3.6. For all observers except observer 2, no statistically significant difference was found between real and simulated lesions (all 95% CI included the value 0.5, p> 0.05). Their AUCs varied from 0.49 to 0.61. Observer 2 had an AUC of 0.63 (95% CI [0.51, 0.74]) indicating that he could differentiate between real and simulated lesions. Observer 5 was the most experienced radiologist while observer 4 was the least experienced radiologist with AUCs 0.54 and 0.51 respectively. The average of all readers' AUC was 0.55 (95% CI [0.44, 0.66], p>0.05). Since the 95% CI includes the value 0.5 (p>0.05), the NH cannot be rejected. In other words the simulated models could not be distinguished from the real microcalcification clusters.

Table 3.3 Area under the ROC curve (AUC) along with the 95% CI for the 2D study per observer; last row reports the average AUC over all readers and the corresponding 95% CI. An asterisk (*) indicates statistical significant difference between real and simulated lesions.

Observer	**Area under the ROC curve (AUC)**	
	AUC	95% CI
1	0.61	[0.49, 0.72]
2*	0.63	[0.51, 0.74]*
3	0.54	[0.42, 0.65]
4	0.51	[0.39, 0.63]
5	0.54	[0.41, 0.65]
6	0.49	[0.36, 0.61]
Average	0.55	[0.44, 0.66]

Some examples of simulated and real cases in the 2D FFDM study are shown in Figs. 3.7 and 3.8 respectively. Figure 3.7 shows false positive cases, i.e. simulated cases interpreted as real cases (threshold point < 3) by at least three radiologists. The real clusters in Fig. 3.8 are false negative cases, i.e. real cases interpreted as simulated (threshold point > 3) by at least three radiologists.

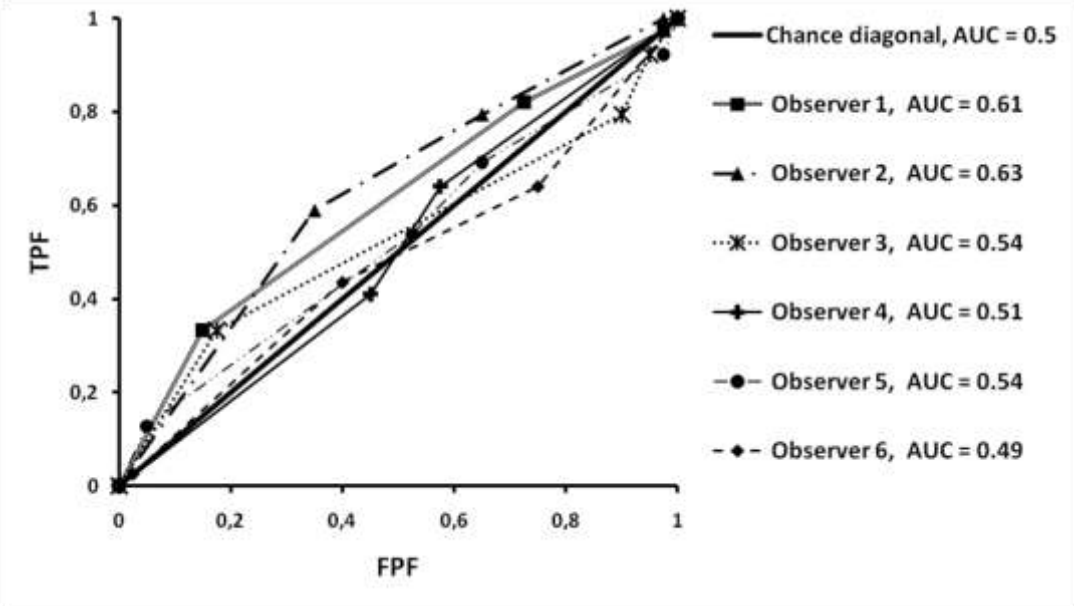
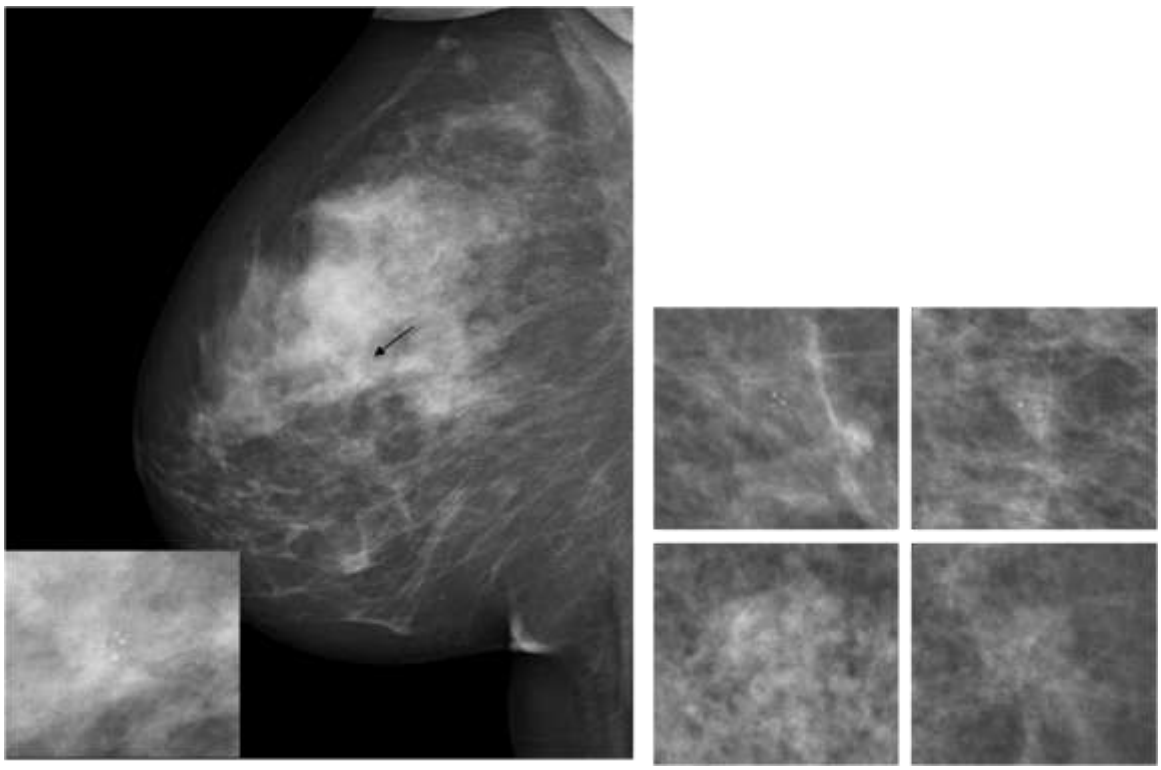
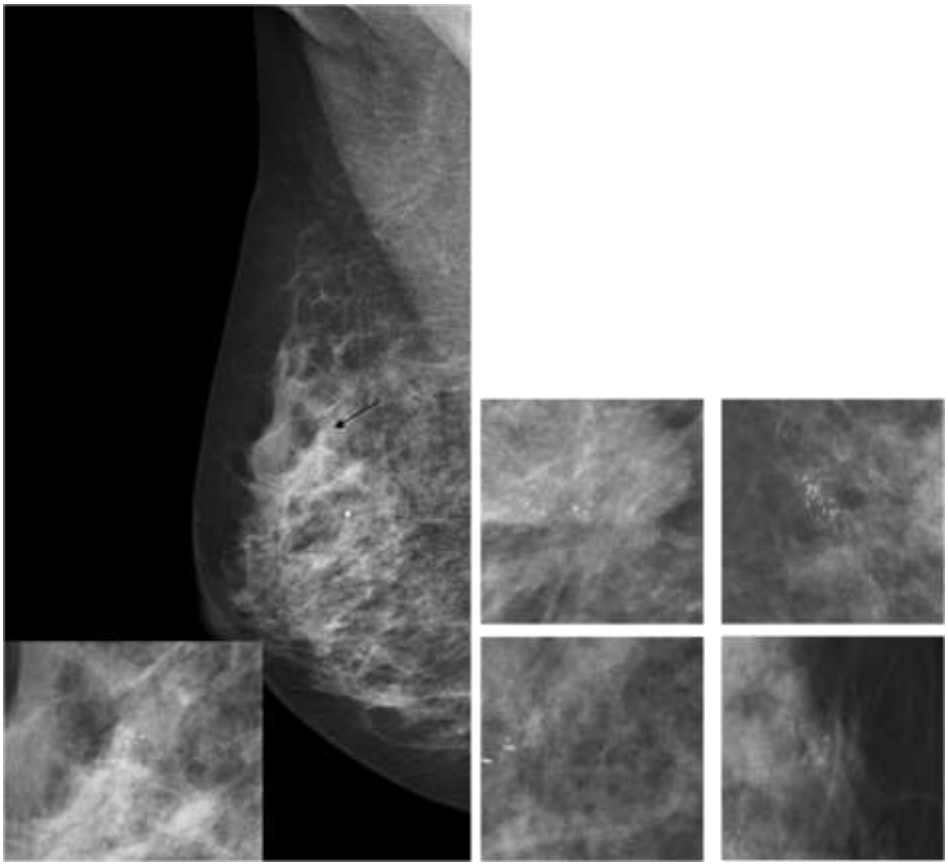


Fig. 3.6. ROC analysis of the 2D FFDM study with area under the ROC curve (AUC) per observer. The false positive fraction (FPF) is plotted against the true positive fraction (TPF). The AUC is calculated using the trapezoidal method.



a) b)

Fig. 3.7. Examples of false positive cases in the 2D FFDM study, i.e. simulated microcalcification clusters thought to be real by at least three radiologists. a) The simulated cluster is shown within the processed mammogram and highlighted with the arrow, in the left bottom corner a magnification view of the cluster. b) Other false positive cases in magnification view.



a) b)

Fig. 3.8. Examples of false negative cases in the 2D FFDM study, i.e. real microcalcification clusters thought to be simulated by at least three radiologists. a) A processed mammogram with a real cluster highlighted with the arrow, on the left bottom corner a magnification of this cluster. b) Other false negative clusters in magnification view.

3.3.2. Results of the tomosynthesis study

Results of the ROC analysis per observer are shown in Table 3.4 and Fig. 3.9. For all observers except observer 1, no statistically significant difference was found between real and simulated lesions (all 95% CI included the value 0.5, $p > 0.05$). The AUCs varied from 0.33 to 0.55. Observer 1 had an AUC of 0.33 (95% CI [0.22, 0.44]) indicating that the observer called most of the simulated lesions to be real and the real to be simulated. Observer 5 was the most experienced radiologist while observer 4 was the least experienced radiologist with AUCs 0.38 and 0.50 respectively. The average of all readers' AUC was 0.46 (95% CI [0.29, 0.64], $p>0.05$) indicating no statistically significant difference between simulated and real clusters.

Table 3.4. Area under the ROC curve (AUC) along with the 95% CI for the tomosynthesis study per observer; last row reports the average AUC over all readers and corresponding 95% CI. Observer 5 was the most experienced radiologist while observer 4 was the least experienced radiologist. An asterisk (*) indicates statistical significant difference between real and simulated lesions.

Observer	**Area under the ROC curve (AUC)**	
	AUC	95% CI
1*	0.33	[0.22, 0.44]*
2	0.51	[0.38, 0.63]
3	0.53	[0.43, 0.62]
4	0.55	[0.43, 0.66]
5	0.38	[0.26, 0.50]
6	0.50	[0.38, 0.62]
Average	0.46	[0.29, 0.64]

Some examples of simulated and real cases participating in the tomosynthesis study are shown in Figs. 3.10 and 3.11 respectively. Figures 3.10 (a) and 3.11 (a) show the in-focus reconstructed plane of the whole breast, with a zooming-in on the cluster in question, while Figs. 3.10 (b) and 3.11 (b) show other examples of simulated and real clusters of microcalcifications in magnification view. The in-focus plane was shown but most of the microcalcifications were distributed over multiple planes within the same cluster. The window level was adjusted for every image for better visualization. The simulated clusters in Fig. 3.10 are false positive cases, i.e. simulated cases interpreted as real cases (threshold point < 3) by at least three radiologists. The real clusters in Fig. 3.11 are false negative cases, i.e. real cases interpreted as simulated (threshold point > 4) by at least three radiologists.

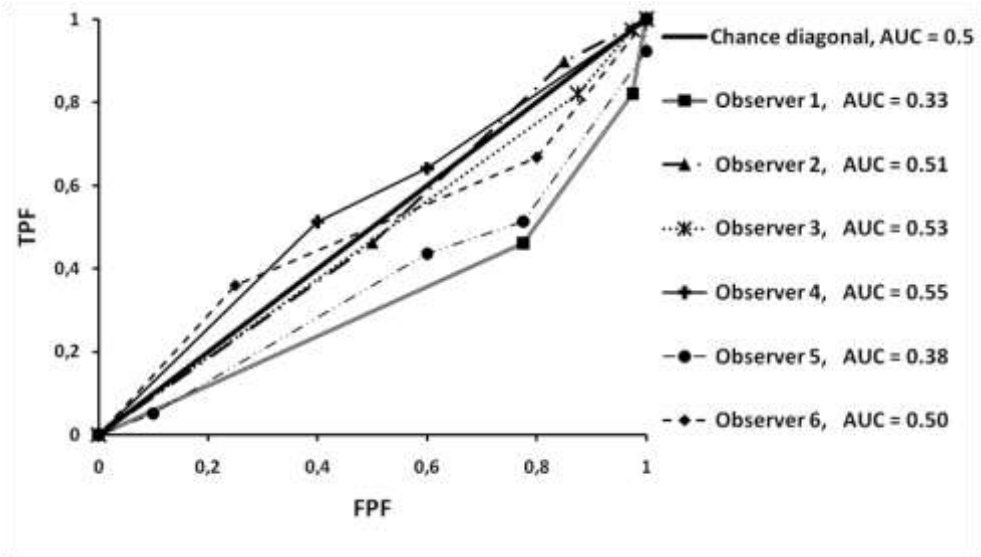
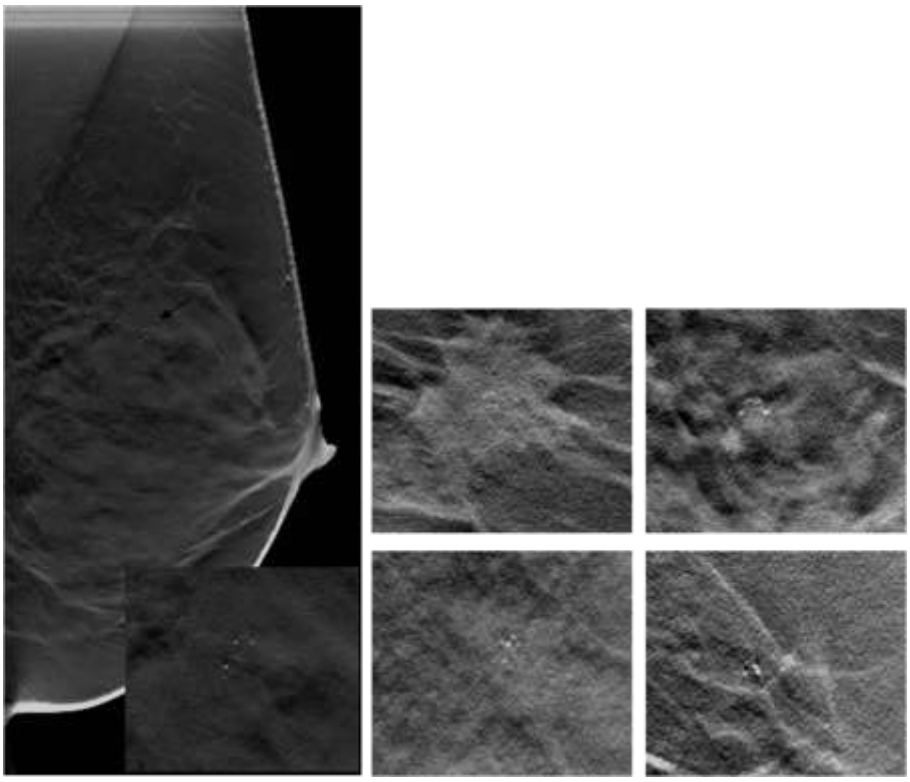
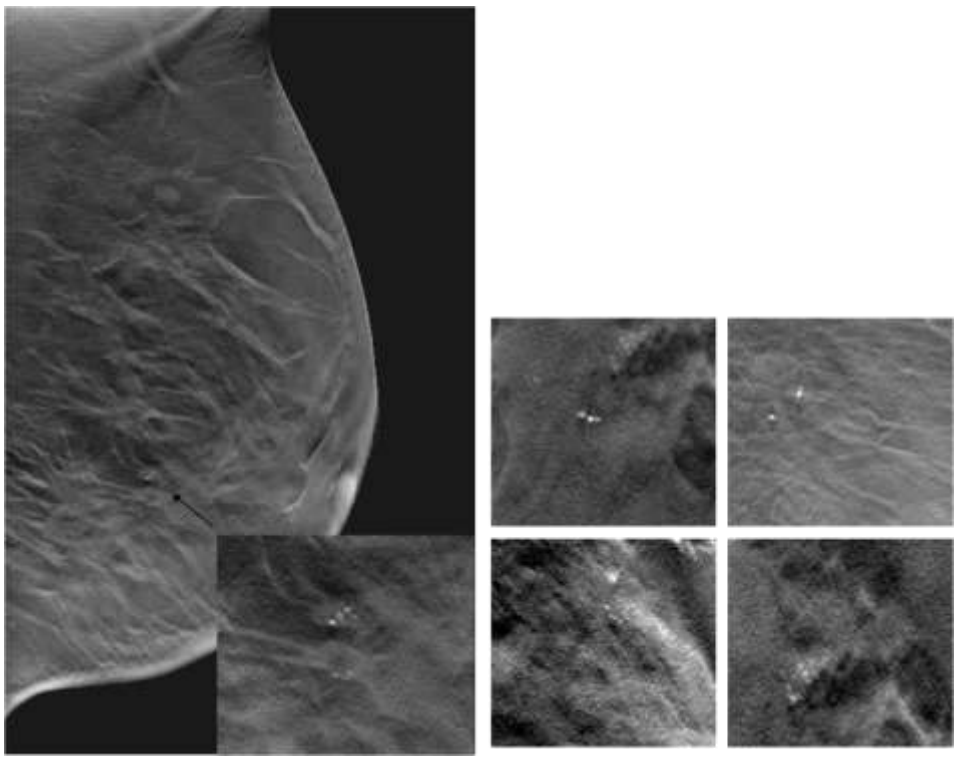


Fig. 3.9. ROC analysis of the tomosynthesis study with area under the ROC curve (AUC) per observer. The false positive fraction (FPF) is plotted against the true positive fraction (TPF). The AUC is calculated using the trapezoidal method.



a) b)

Fig. 3.10. Examples of false positive cases (i.e. simulated microcalcification clusters thought to be real cases by at least three radiologists) in the tomosynthesis study with the in-focus plane presented, while the microcalcifications were distributed over multiple planes within the cluster. The window level was changed for every image for better visualization. a) The simulated cluster is shown within the breast with the arrow pointing to it, in the right bottom part the cluster in magnification view. b) Other false positive cases in magnification view.



a) b)

Fig. 3.11. Examples of false negative cases (i.e. real microcalcification clusters thought to be simulated by at least three radiologists) in the tomosynthesis study with only the in-focus plane presented, but most of the microcalcifications were distributed over multiple planes within the same cluster. The window level was modified for every image for better visualization. a) The real cluster is shown within the breast with the arrow pointing to it, in the right bottom part the cluster in magnification view. b) Other false negative cases in magnification view.

3.3.3. Comparison between 2D FFDM and breast tomosynthesis

Table 3.5 shows the results of the kappa statistical analysis. The agreement of the readers for the group with simulated clusters ranged from -0.12 to 0.21 which was interpreted as “poor” to “fair” agreement with an average of 0.06 indicating “poor” agreement. As for the group with real clusters, it ranged from -0.15 to 0.14. This is “poor” agreement with an average of 0.06 identical to the value of the group with simulated clusters and categorized as “poor” agreement.

Table 3.5. The agreement, as measured by the kappa statistic, between the 2D FFDM and the tomosynthesis studies for the simulated population and the real population. Listed are the values of the kappa statistic, the 95% confidence interval, and the conventional interpretations of the kappa values for each observer and the average.

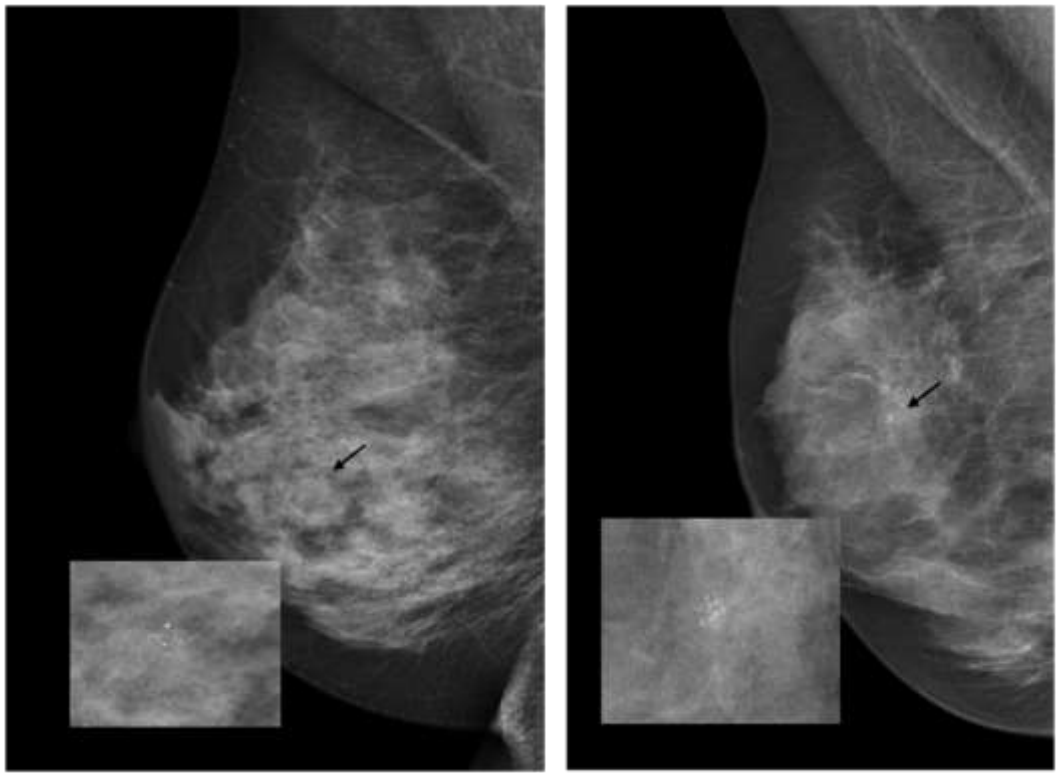
Observer	Simulated		Real	
	kappa, (95% CI)	Interpretation	kappa, (95% CI)	Interpretation
1	0.18 (-0.12,0.48)	Poor agreement	-0.15 (-0.27,-0.03)	Poor agreement
2	0.21 (-0.01,0.43)	Fair agreement	0.11 (-0.12,0.34)	Poor agreement
3	-0.06 (-0.12,-0.01)	Poor agreement	0.05 (-0.13,0.23)	Poor agreement
4	0.20 (-0.07,0.47)	Poor agreement	0.06 (-0.21,0.33)	Poor agreement
5	-0.06 (-0.28,0.16)	Poor agreement	0.14 (-0.08,0.36)	Poor agreement
6	-0.12 (-0.34,0.10)	Poor agreement	0.14 (-0.11,0.38)	Poor agreement
Average	0.06 (-0.25,0.37)	Poor agreement	0.06 (-0.16,0.28)	Poor agreement

3.4. Discussion

Interest in the development of 3D models of breast lesions [7, 29, 31, 61] is increasing rapidly due to the introduction of 3D systems such as DBT and BCT. However, most of the currently developed models are rather simplified versions of real lesions that are either embedded in homogeneous backgrounds, synthetic (clustered lumpy) backgrounds [61], and anthropomorphic breast models [62]. These studies usually follow a simulation approach either based on Monte Carlo simulations or a full simulation approach such as that described by Gong et al. [29]. Whether the use of simplified models is justified depends upon the type of task to be studied. Optimization studies that aim for an improved detection of particular lesions require more realistic models. Therefore, some observer studies have used simulated lesions in patient images. The simulation of masses was the main focus of these studies [31]. As the detectability and characterization of microcalcifications is an issue that requires further optimization in DBT systems, 3D models of microcalcifications might prove extremely useful. The currently proposed 3D models of microcalcifications are usually simple shapes represented by small spheres distributed in regular patterns or randomly distributed [7]. In both cases, these models represent only a small fraction of the variety of shapes of microcalcifications and the different distributions of these microcalcifications within the cluster. Some other morphological shapes include irregular forms or linear structures. Segmented 2D FFDM models have been proposed and validated [48], but their extension to 3D models is not straightforward. The present paper focused on the development of 3D models of microcalcification clusters. When extracted from real biopsy specimens containing clusters of microcalcifications, the exact morphology and histological diagnosis are

known. In this study, we did not select particular types of biopsy specimen: the selection was based upon the access to the biopsy specimen for a reasonable period of time, the presence of all scientists in charge of the work and a time slot available on the micro-CT scanner. In this regard, our set of models is representative of the variation of microcalcification clusters biopsied at our clinical centre and is therefore a clinically relevant sample. The database is currently based on 23 biopsies, but has been enlarged to 54 3D models by rotating a number of clusters.

The 3D models of microcalcifications were validated for two modalities, 2D FFDM digital mammography and breast tomosynthesis, and the results indicated that the proposed method can produce simulated microcalcification clusters for both of these modalities that cannot be distinguished from real clusters. In order to gain additional insight and improve future simulation studies, the evaluations were discussed amongst the radiologists once the observer study was finished. Overall, the radiologists felt that the real and simulated clusters could not be distinguished, but they had greater difficulty with the tomosynthesis study compared to the 2D FFDM study due to their greater experience with 2D FFDM images. Some of the difficulties with the tomosynthesis study included an apparently higher amount of noise in the tomosynthesis images when compared to 2D images. In addition, the presence of out of plane artifacts and the extension of the microcalcifications over several planes were all newly introduced facts which were important especially when the whole cluster had to be judged. Some true negative cases in the 2D FFDM study, i.e. simulated cases where at least three radiologists agreed that they were simulated (threshold point > 3), were selected for more in depth discussion. Some of the simulated clusters had higher contrast than expected, some clusters looked artificial when magnified, and when looking at the whole breast, a few clusters looked super-imposed and not embedded in the surrounding tissue. These points are illustrated in Fig. 3.12. The cluster in Fig. 3.12 (a) had high contrast but the morphology was considered realistic. The cluster in Fig. 3.12 (b) was accepted in terms of morphology, however when enlarged (magnified) the microcalcifications looked artificial compared to other structures in the breast and it was also considered super-imposed.



a) b)

Fig. 3.12. Some examples of true negative cases, i.e. simulated microcalcification clusters where at least three radiologists agreed that they were simulated, in the 2D FFDM study. a) The cluster had high contrast. b) The cluster looked superimposed and artificial when magnified.

The agreement test, using the kappa statistic, was implemented mainly to check if there is an agreement between the scoring of the radiologists for the two modalities. It was shown that there is on average "poor" agreement between each reader judging the realism of the appearance of the same cluster in 2D FFDM and tomosynthesis. The test was divided into separate tests, for the simulated and real populations to further investigate the effect of the simulation compared to real microcalcification clusters, i.e. check whether this disagreement for the whole population was reflected in both populations separately or not. The results demonstrated that this "poor" agreement is common between both populations, an indication that the simulation is realistic and will not bias the judgment of radiologists when used in observer comparative studies. This result suggests the possibility of using these models for comparative studies between 2D mammography and breast tomosynthesis.

The presented simulation framework has applications such as comparative observer performance studies in terms of detectability of lesions between 2D FFDM, breast tomosynthesis, and presumably breast computer tomography and to compare different parameters in the same modality in order to

optimize system performance such as reconstruction algorithms. It can also be used to guide phantom design for quality assurance tasks. The framework as it stands can be considered reasonably general as lesions can be simulated into projection images for any system for which the geometrical information is known and the system MTF can be established. This would have to be verified first. Although different acquisition settings such as tube voltage and anode/cathode filter can be accommodated, these will just change the lesion contrast and not the contrast of the background structures or dynamic range present in the imaged breast (set by the original acquisition). Noise is inherited from the background and hence full dose optimization studies cannot be conducted using this framework, although it may be possible to add noise, simulating the effect of dose reduction. A possible solution to this problem is the use of physical phantoms that generate backgrounds with a spatial frequency content consistent with various anatomies, as suggested by Gang et al. [63]. The lesions could be simulated into reasonably realistic anatomical backgrounds at the chosen dose. Alternatively, the 3D models could be used in different simulation approaches, whether by their inclusion into voxel phantoms representing the breast [62, 64–73] or simulated clustered lumpy backgrounds [29, 61, 74].

Although the simulation of the 3D microcalcification clusters was successfully validated, some limitations were present in the study. (1) As not all pathology reports for the microcalcification clusters representing the group with real clusters were available, it was not possible to compare the performance of the 3D models in terms of realistic appearance separately for the benign and the malignant clusters. Since most of the simulated clusters were retrieved from malignant lesions, as confirmed by the pathology report, we presume that our conclusions regarding the possibility to simulate clusters realistically are especially valid for this group of lesions. (2) The background characteristics around the simulated lesions were not changed. This issue is of high importance for the simulation of malignant masses with spiculated or irregularly shaped borders due to the invasion of the tumour in the surrounding breast tissue. Microcalcifications will not distort the background tissue because their sizes are rather small. We hypothesized that the most important aspects to realistically simulate malignant calcifications with irregular borders are the accuracy of the 3D model that captures the malignant/benign characteristics and the position of insertion into the breast. These aspects were considered in the simulation procedure. (3) Calcium oxalate was chosen as a substitute material to simulate microcalcifications. In the literature, several materials were also used to represent microcalcifications such as calcium carbonate [61], calcium phosphate [7] and calcium hydroxyapatite [75]. Fandos-Morera et al. [52] report that calcium oxalate was found in malignant microcalcifications while apatite was found in malignant and benign microcalcifications. Haka et al.

[76] reported that calcium oxalate was found in benign calcifications while hydroxyapatite (which is a type of apatite) was found in malignant calcifications. This introduces some uncertainty into the exact composition of microcalcification material. In reality some clusters are composed of malignant and benign calcifications that cannot be differentiated. Therefore, type specific values could not be attributed. The choice of calcium oxalate in comparison to apatite was made based on the linear attenuation coefficient of these materials. For example, a microcalcification with peak thickness 0.1 mm at mean energy 20 keV would have as contrast: 0.146 for apatite (ρ=3.19 g/cm3, (μ/ρ)=4.95 cm2/g) and 0.092 for calcium oxalate (ρ=2.12 g/cm3, (μ/ρ)=4.58 cm2/g). From few preliminary tests, it has been seen that the lower contrast gave a more realistic appearance to the simulated clusters. A more detailed study that examines how contrast values are propagated into processed 2D digital mammograms and breast tomosynthesis images would be interesting future work. (4) Final limitations are related with the adjustment of the templates in our simulation framework. We currently take detector MTF, the oblique entry of the beam on the detector and focal spot motion as apparent at the level of the detector into account. Extra focal spot motion induced blurring for objects away from the detector and towards the x-ray source (the z-position in the simulation) is not yet included and will most likely improve the contrasts of simulated calcification.

3.5. Conclusion

A database of 3D models of microcalcification clusters has been created and the realistic appearance of the models validated; a ROC analysis was performed on 40 simulated and 40 real clusters in 2D mammograms and in breast tomosynthesis series. The results revealed that observers could not distinguish between real and simulated clusters based on the area under the ROC curves averaged over all observers. The AUC was 0.55 (95% confidence interval [0.44, 0.66]) for the 2D study, and 0.46 (95% confidence interval [0.29, 0.64]) for the tomosynthesis study. The “poor” agreement based on the kappa statistic test confirmed this conclusion. Following validation of the realistic appearance of these models for the two modalities (2D FFDM and DBT), this database can now be used in future observer performance studies to optimize system parameters in both 2D FFDM digital mammography and breast tomosynthesis.

ACKNOWLEDGMENTS

This work is part of the OPTIMAM project which is funded by CR-UK & EPSRC Cancer Imaging Program in Surrey, in association with the MRC and Department of Health (England). We are grateful to the radiologists who participated in the studies: Dr. Andreas Van Steen, Dr. Sandra Postema, Dr.

Kirsten Joossens, Dr. Ilse Vervloessem and Dr. Riet D'Hauwe. We would like to thank Thomas Mertelmeier from Siemens (Erlangen, Germany) for providing the reconstruction software "TomoEngine" and the reprocessing software "OpView2".

Chapter 3

Chapter 4

The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis

adapted from E. Shaheen et al., Med. Phys. 2014, accepted with minor revisions

Chapter 4

Abstract

Purpose: This work proposes a new method of building 3D breast mass models with different morphological shapes and describes the validation of the realism of their appearance after simulation into 2D digital mammograms and breast tomosynthesis images.

Methods: Twenty-five contrast enhanced MRI breast lesions were collected and each mass was manually segmented in the three orthogonal views: sagittal, coronal and transversal. The segmented models were combined, resampled to have isotropic voxel sizes, triangularly meshed and scaled to different sizes. These masses were referred to as non-spiculated masses and were then used as nuclei onto which spicules were grown with an iterative branching algorithm forming a total of 30 spiculated masses. These 55 mass models were projected into 2D projection images to obtain mammograms and into tomographic sequences of projection images, which were then reconstructed to form 3D tomosynthesis datasets. The realism of the appearance of these mass models was assessed by five radiologists via receiver operating characteristic (ROC) analysis when compared to 54 real masses. All lesions were also given a BIRADS score. The data sets of 2D mammography and tomosynthesis were read separately. The Kendall's coefficient of concordance was used for the inter-rater observer agreement assessment for the BIRADS scores per modality. Further paired analysis, using the Wilcoxon signed rank test, of the BIRADS assessment between 2D and tomosynthesis was separately performed for the real masses and for the simulated masses.

Results: The area under the ROC curves, averaged over all observers, was 0.54 (95% confidence interval [0.50, 0.66]) for the 2D study, and 0.67 (95% confidence interval [0.55, 0.79]) for the tomosynthesis study. According to the BIRADS scores, the non-spiculated and the spiculated masses varied in their degrees of malignancy from normal (BIRADS 1) to highly suggestive for malignancy (BIRADS 5) indicating that we had managed to create different types of models. The assessment of the BIRADS scores for all observers indicated good agreement based on Kendall's coefficient for both the 2D and the tomosynthesis evaluations. The paired analysis of the BIRADS scores between 2D and tomosynthesis for each observer revealed consistent behaviour for the real and simulated masses.

Conclusions: A database of 3D mass models, with variety of shapes and margins, was validated for the realism of their appearance for 2D digital mammography and for breast tomosynthesis. This database is suitable for use in future observer performance studies whether in virtual clinical trials or in patient images with simulated lesions.

4.1. Introduction

In the previous chapter, a database of 3D models of microcalcification clusters with different morphological shapes was presented and validated for realistic appearance in two dimensional full field digital mammography (2D FFDM) and DBT [34]. In this chapter, we focus on masses. Mass models for both 2D and 3D imaging are being developed. To simulate 2D mass models, real masses from clinical patient images have been characterized and subsequently synthesized by various authors [51, 59, 77, 78]. These authors validated the realism of the appearance of their models by testing whether radiologists could distinguish between real and simulated masses when embedded in 2D patient images. Ruschin *et al.* [79] produced 2D masses with different shapes such as oval (based on the method proposed by Skiadopoulos *et al.* [59]), irregular margins based on random walk and spiculated masses by adding star-shaped objects to a nucleus. The use of these 2D models is obviously limited to simulations in 2D images. An equivalent approach for 3D modalities therefore requires the development of 3D models of masses. Bliznakova *et al.* [62] presented 3D mass models as part of their 3D breast phantom in two forms: simple circumscribed models simulated by ellipsoids or spheres and irregularly shaped models based on random walk. Gong *et al.* [29] and Hinstala *et al.* [80] adopted the same concept of random walk to simulate 3D irregularly shaped masses. Ruschin *et al.* [81] approached the same task by using tumour shapes from 2D simulations [79], with the original 2D model as the central plane in the 3D object and the planes above and below were smaller replicas of the central plane. Attempts to create 3D spiculated mass models were limited to two methods: 1. 2D spiculated images were hand drawn and connected using a non-uniform rational B-spline (NURBS) surface [7]. 2. Branches were grown iteratively on a Gaussian random sphere model [82]. As far as we know, the realism of the appearance of these 3D models has not yet been validated in observer studies even though they have been used in virtual clinical studies. More recently, Rashidnasab *et al.* [83] developed 3D mass models based on diffusion limited aggregation. These masses have irregular shapes (without spicules) and were proven to be more realistic than models obtained with approaches using random walk algorithms when projected and validated in a 2D study.

The number of validated 3D mass models with different shapes available for further use in simulation work is very limited. Therefore, the purpose of this study was to create clinically-relevant 3D models with a variety of morphological shapes and to validate the realism of their appearance for future use in simulation studies. Two methods have been developed. The first generates masses that appear either benign or suspicious with an intermediate degree of malignancy. The second method is an improved implementation for the development of more malignant breast masses. All models have

been carefully validated in terms of the realism of their appearance in both 2D and DBT patient images.

4.2. Materials and Methods

The simulation of mass models started with the collection of breast MRI lesions that were segmented and scaled. These represented non-spiculated masses (method 1). The same models were then used as a nucleus onto which branches (spicules) have been grown (method 2). A set of 3D mass models was generated and the validated simulation framework [33], presented in chapter 2, was applied to these models to simulate them into projection images. The realism of their appearance was then validated for 2D FFDM and DBT.

4.2.2. Simulation of 3D mass models

According to the BIRADS lexicon [84], a "Mass" is defined as a three-dimensional structure demonstrating convex outward borders, usually evident on two orthogonal views. Masses can be classified by shape, margin and density. The shape of a mass can be round, oval, lobular or irregular. The margin can be described as circumscribed (well-defined or sharply-defined), indistinct (ill-defined) or spiculated. When compared to adipose tissue, the density of a mass can be higher, lower or equivalent [84, 85]. To form a database with a clinically representative variety of morphological shapes, we developed methods to simulate all possible shapes and margins.

4.2.2.1. Non-spiculated 3D mass models

Breast MRI acquisitions of twenty histologically-proven malignant masses and five benign cases, imaged with a 3D contrast enhanced acquisition, were collected. The use of anonymized images is allowed by the ethical committee of the hospital. Each mass was manually segmented in three reconstructed orthogonal planes (sagittal, transversal, coronal) and the resulting segmentations were combined with a logical OR. An example of a malignant lesion in the three orthogonal planes and the corresponding voxelized 3D model is shown in Fig. 4.1. The voxel size of the selected cases was 1.042x1.042x2.0 mm^3. The 3D model was resampled to have isotropic voxel sizes of 1.042 mm in 3D space. The models were then meshed and scaled to four different resulting diameter sizes of their bounding sphere: very small (diameter ~75 voxels), small (~100 voxels), medium (~150 voxels) and large (~250 voxels). The models were voxelized and saved in binary format with 1 for the mass and 0 for background. The first three sizes were used for further modifications as described in section

4.2.2.2 and will be referred to as “central masses”. The large size lesions, will be referred to as “non-spiculated masses”, and were simulated into 2D and DBT images using the simulation framework. The segmented masses had different shapes including round, oval, lobular and irregular. The segmentation steps for this part were implemented using MeVisLab (version 2.3.1, MeVis Medical Solutions AG, Bremen, Germany).

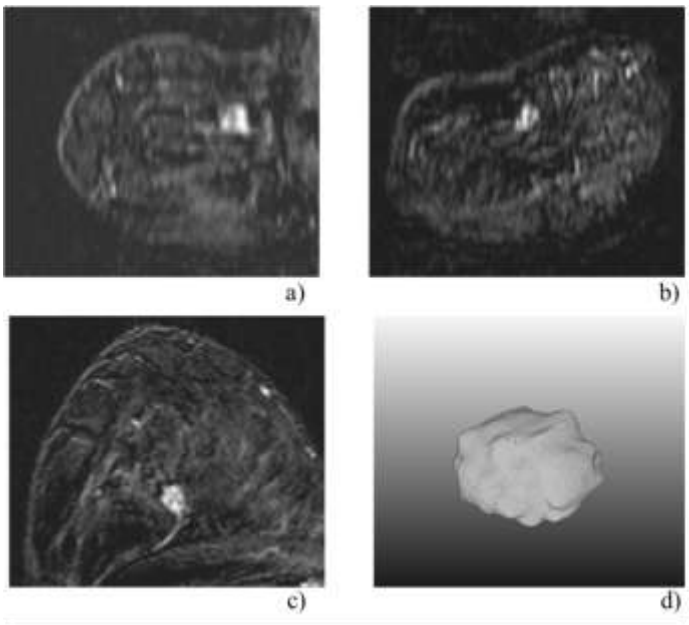


Fig. 4.1. An example of an MRI lesion in: a) sagittal view, b) coronal view, c) transversal view, d) the 3D voxelized model of the same mass.

4.2.2.2. Spiculated 3D mass models

Due to the low resolution of MRI images, spicules were not captured in the segmentation procedure. As a result, these masses had either circumscribed or indistinct margins. In order to create spiculated masses, with higher suspicion for malignancy, branches were grown on the surface of each central mass following an iterative branching algorithm [82]. The spicules were created by iteratively adding segments with decreasing lengths and diameters. The algorithm was divided into two parts: 1. central mass preparation, 2. iterative branching algorithm.

1. Central mass preparation

The central mass was triangularly meshed and the number of triangles was reduced to allow different base sizes of branches: thin, medium and thick. The central mass was divided into three regions along

the z-axis: upper, middle and lower regions (Fig. 4.2). The branches grown on the upper and lower surface regions were bent outward to avoid artificially looking white spots in the projection images originating from spicules that are parallel to the z-axis. For every region and for every triangle, the in-centre of the triangle, the radius of the in-circle and the orientation of the normal to the surface of the triangle were calculated (Fig. 4.2). These parameters were used as input to the iterative branching algorithm to grow a branch with the in-circle radius as base to the first segment, at the in-centre 3D position.

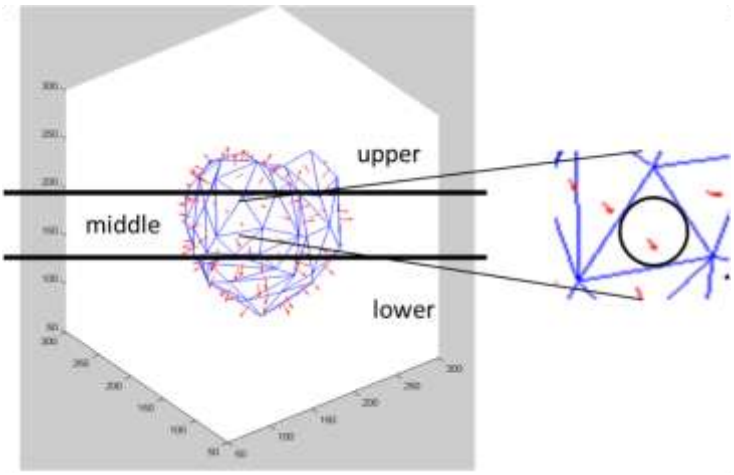


Fig. 4.2. A central mass in the preparation step: meshed triangularly, divided into upper, middle and lower regions in the Z-direction. For every triangle, the in-circle radius, the normal to the triangle surface (red arrow) located at the in-center were calculated.

2. Iterative branching algorithm

The branch is simulated as a tubular shape consisting of a set of segments attached on top of each other. In our implementation, a segment was represented by a circular oblique cut cone defined by: radius of the bottom (r_b), radius of the top (r_t), length of the cut cone (l), inclination angle (ϕ) (i.e. the angle from the z-axis towards the y-axis) and the rotation angle (θ) (i.e. the angle around the z-axis from the x-axis) as shown in Fig. 4.3.

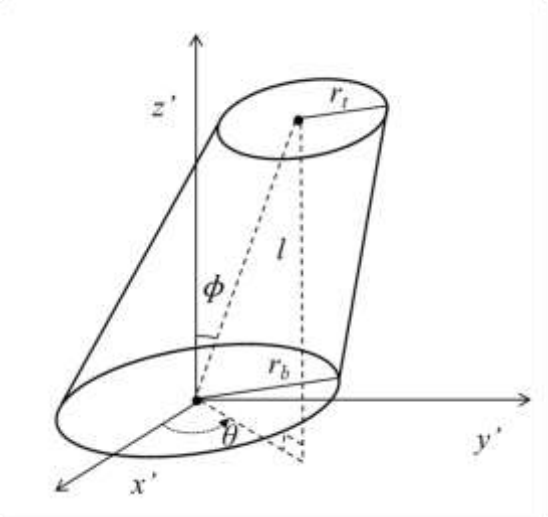


Fig. 4.3. The geometry of the circular oblique cut cone and the definition of the parameters: radius of the bottom (r_b), radius of the top (r_t), length of the cut cone (l), inclination angle (ϕ) from z-axis and rotation angle (θ) around z-axis.

The basic growing rules of the algorithm [82] for each iteration are:

- Each segment will produce zero, one or two children, with probabilities randomly selected.
- The children have a starting point equal to the end point of the parent, an initial radius smaller than the final radius of the parent and a direction of growth slightly different from that of the parent.
- In case of a bifurcation, the child segments will be scaled in diameter similar to the branches found in the arterial system. The scaling is based on the principle of minimum work (PMW) as proposed by Murray [86]. The PMW is a parametric optimization model for the growth and adaptation of arterial trees. According to this theory, the optimum rule for connecting arteries is that at branching points the cube of the radius of a parent artery is equal to the sum of the cubes of its branches [86]. An indirect measure of the division of flow at branching points is the diameter exponent α (Eq. 4.1) [87], which was proposed by Murray to be equal to 3 in-line with the above theory. Other studies [88] showed that this theoretical value slightly overestimated the truth and they performed measurements and simulations with α ranging between 2 and 4 yielding an optimum value of α =2.7.

$$r_0^{\alpha} = r_1^{\alpha} + r_2^{\alpha} \tag{4.1}$$

Consider the two extreme situations in case of a bifurcation with the assumption that r_1 is the larger of the two children radii:

1. For a bifurcation with equal-sized children branches ($r_1 = r_2$), it follows from Eq. 4.1 that

$$r_1^{\alpha} = \boldsymbol{0.5}\, r_0^{\alpha} \tag{4.2}$$

2. For a bifurcation with one very small branch ($r_2 \approx 0$), Eq. 4.1 simplifies to

$$r_1^{\alpha} = \boldsymbol{1}\, r_0^{\alpha} \tag{4.3}$$

A variable was introduced by De Sisternes *et al.* [82], called the dividing ratio (*d*), that ranges from 0.5 to 1 in Eq. 4.4 (still assuming that r_1 is the larger of the two children branches)

$$d = \frac{r_1^{\alpha}}{r_0^{\alpha}} \tag{4.4}$$

Combining Eq. 4.4 and Eq. 4.1, we get:

$$r_1 = r_0\, d^{1/\alpha} \tag{4.5}$$

$$r_2 = r_0\, (1-d)^{1/\alpha} \tag{4.6}$$

At each iteration, each parent will produce one of four possible options:

1. *No child:* No segment is added when the radius or length of the parent reaches zero.
2. *One child (continuing the branch):* The rules defined earlier will be applied on the child segment input parameters as follows:
   - The radius of the bottom of the child segment is equal to the radius of the top of the parent segment
   - The radius of the top of the child is reduced slightly following Eq. 4.7:

$$r_t = r_b \cdot (dr_r)^{\frac{1}{2.7}} \tag{4.7}$$

     Where dr_r is a factor defining how fast the radius decreases, and ranges between 0 and 1. The diameter exponent was chosen to be 2.7 [88].
   - The length of the child is reduced slightly following Eq. 4.8:

$$l_{child} = l_{parent} \cdot dr_l \tag{4.8}$$

     Where dr_l is a factor defining how fast the length decreases.
   - The orientation of the child segment is determined by the rotation angle (θ) and the inclination angle (ϕ).

$$\theta_{child} = \theta_{parent} + \gamma_{bif\_theta} \quad (4.9)$$

$$\Phi_{child} = \Phi_{parent} + \gamma_{bif\_phi} \quad (4.10)$$

Where γ_{bif_theta} and γ_{bif_phi} are the bifurcation angle factors for both the rotation and inclination angles respectively. They are selected by the user and follow a normal distribution with values ranging between 0 and π tending to smaller values.

3. *Two children:* there are two options with two children, either symmetric bifurcation or asymmetric bifurcation. The rules for one child can be applied with exception for the orientation, where the children grow in different directions, and their radii scale following the dividing ratio (*d*) introduced in eq. 4. In this study, *d* was set to 0.5 in case of symmetric bifurcation and 0.8 for asymmetric bifurcation. The new equations for the scaling of radii, lengths and growing orientations for the children are as follows:

$$r_{b1} = r_b \cdot (dr_r \cdot d)^{\frac{1}{2.7}} \quad (4.11)$$

$$r_{b2} = r_b \cdot \left(dr_r \cdot (1-d)\right)^{\frac{1}{2.7}} \quad (4.12)$$

$$l_1 = l_{parent} \cdot dr_l \cdot d \quad (4.13)$$

$$l_2 = l_{parent} \cdot dr_l \cdot (1-d) \quad (4.14)$$

$$\theta_{child1} = \theta_{parent} + \gamma_{bif\_theta} \quad (4.15)$$

$$\theta_{child2} = \theta_{parent} + \pi - \gamma_{bif\_theta} \quad (4.16)$$

$$\Phi_{child1} = \Phi_{parent} + \gamma_{bif\_alpha} \quad (4.17)$$

$$\Phi_{child2} = \Phi_{parent} + (\gamma_{bif\_alpha} - 1) \quad (4.18)$$

4.2.2.3. Implementation Parameters

1. Non-spiculated masses database

A preliminary study was conducted to filter out unrealistically shaped masses by simulating the 25 mass models into 2D and DBT patient images using the simulation framework. The resulting images were shown to an expert radiologist (radiologist A). The radiologist judged the realism of the appearance of the masses and gave a BIRADS score [84] (Table 4.1). In terms of the realism of appearance, all masses were accepted and the BIRADS scores ranged from 1 to 4 which indicated the presence of different shapes from benign to suspicious for malignancy as shown in Fig. 4.4.

Table 4.1. Breast Imaging Reporting and Data System (BIRADS) categories for mammographic lesions according to their probability of being malignant.

BIRADS	Final Assessment category	Probability of malignancy (%)
0	Need additional imaging evaluation	-
1	Negative	0
2	Benign	0
3	Probably benign	<2
4	Suspicious abnormality	2-95
5	Highly suggestive of malignancy	>95
6	Histologically proven malignancy	100

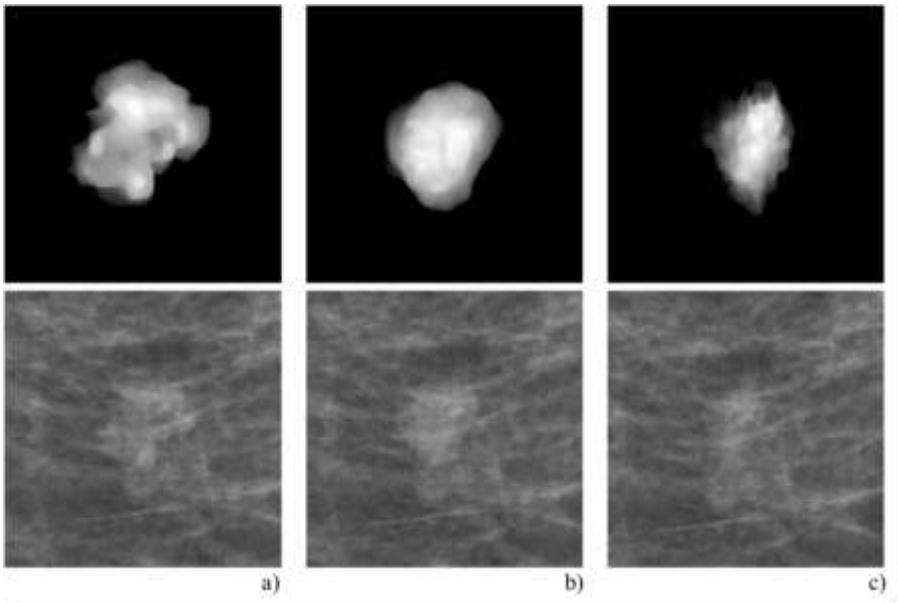


Fig. 4.4. Different shapes and margins of non-spiculated mass models. The top row is the projection image of the 3D models and the bottom row the mass inserted into 2D patient image. The shapes include round/oval (b), lobular (a) and irregular (c). The margins include well-defined (a, b) and ill-defined (c).

2. Spiculated masses database

A number of 3D spiculated masses was developed as described earlier. Based mainly on trial and error, branches were grown on the surface of the 25 central masses following the different designs and more than 80 masses were developed. Different shapes of spiculated masses were grouped into four designs (Table 4.2): general, grouped branches, stellate and hairy (Fig. 4.5). Another preliminary study was conducted in order to test the quality of the masses in terms of the realism of their

appearance. Therefore, each mass was simulated into 2D and DBT images of a patient in the same position. All 80 masses were shown to radiologist A for the scoring of their realism and BIRADS. The expert radiologist accepted only 30 masses and their BIRADS scores ranged from 3 to 5, i.e. the malignant types were included. The eliminated masses were either considered repetitions of other models or unrealistic in terms of the number of branches: some models had too many branches, others too few. A database of 55 spiculated and non-spiculated breast 3D mass models was subsequently built with the BIRADS scores, as judged by radiologist A, ranging from 1 to 5. Validation was performed through comparison to real lesions and other radiologists' assessments as described in section 4.2.4.2.

Table 4.2. The different designs of the spiculated 3D mass models with their descriptions and the central mass sizes used.

Design Name	**Description**	**Central mass size**	**No. of masses**
General	Three sizes of branches were used: thin, medium and thick. Some masses contained either only one of the sizes or combination of different sizes. The branches on top and bottom were bent outward to avoid bright spots when projected	very small, small, medium	15
Grouped	Groups of 20-25 thin branches were attached to the surface of different regions in the middle part of the lesion in the Z direction	small, medium	7
Stellate	Long thin branches randomly distributed in the middle part of the mass in the Z direction	very small	3
Hairy	Short thin branches randomly distributed all over the surface of the mass. The branches on top and bottom were slightly bent outward	small	5

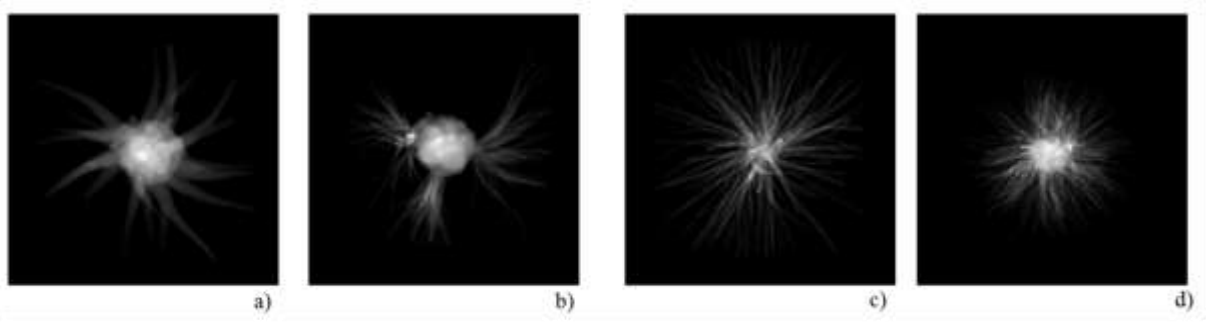


Fig. 4.5. The different designs of spiculated masses presented by their projection images. a) a spiculated mass type "general" with medium and thick branches. b) a spiculated mass type "grouped branches". c) a spiculated mass type "stellate". d) a spiculated mass type "hairy" with only thin branches.

The accepted 30 spiculated masses were based on only 13 different central masses and the sizes used in the simulation per design are shown in Fig. 4.6 and Table 4.2. In order to grow branches with different thicknesses, lengths, orientations and bifurcation options following the iterative branching algorithms, the parameters in equations 4.7 to 4.18 had to be defined previously by the user. Due to the variety of combinations, we report the range of values per parameter in Table 4.3. All the simulation steps to develop spiculated masses were implemented using MATLAB (version 7.8, MathWorks, Natick, MA).

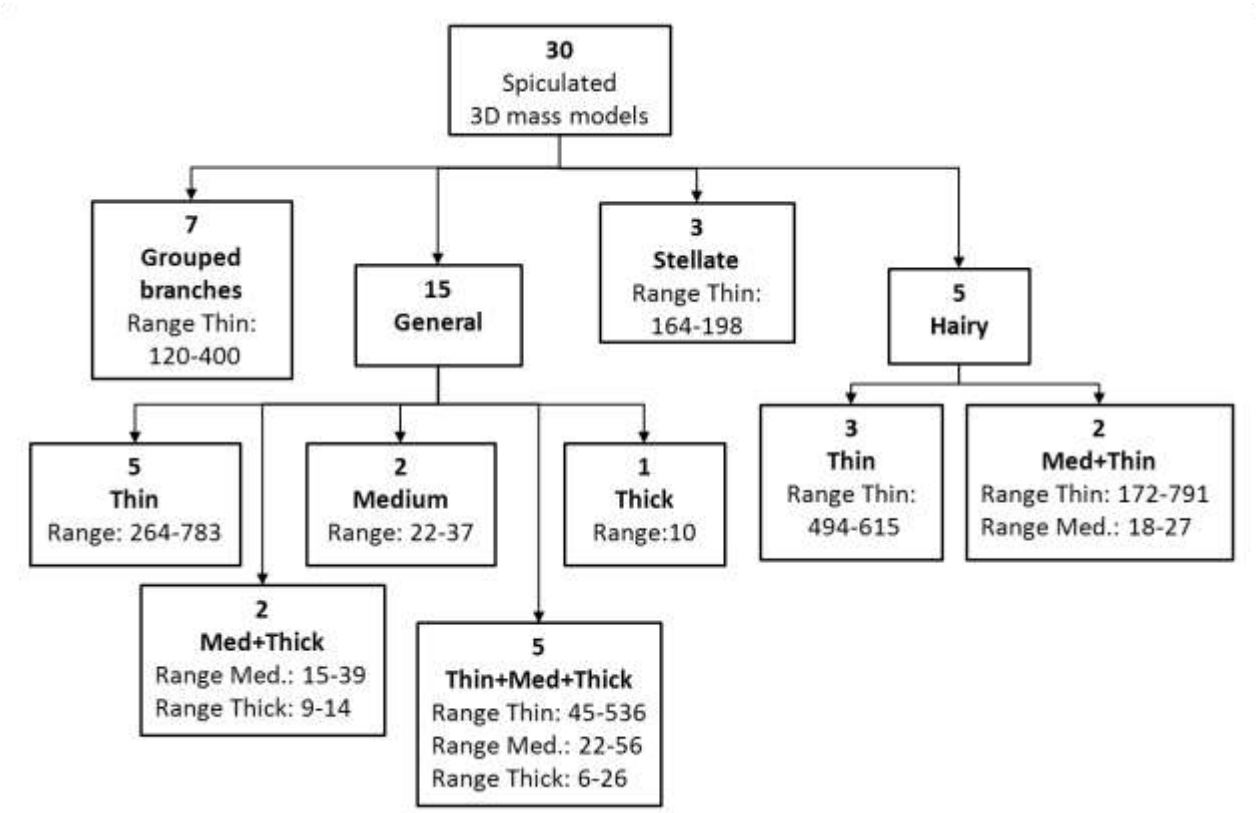


Fig. 4.6. The distribution of the 30 spiculated mass models into the different designs (grouped branches, general, stellate, hairy) and the combinations of the sub-groups which are named after the branch sizes used. The range of the number of branches per size are reported per group/sub-group.

Table 4.3. The range of implemented values for the parameters used to grow branches iteratively (as found in equations 4.7 to 4.18) for the different designs of spiculated masses.

	r_b			dr_r	l_{parent}	dr_l	θ_{parent}	ϕ_{parent}	γ_{bif_theta} γ_{bif_alpha}
	Thin	**Med.**	**Thick**						
Range	2-6 voxels	5-20 voxels	>20 voxels	0.6-0.9	5-30 voxels	0.8-0.95	0°-10°	0°-20°	2°-10°

4.2.3. Simulation framework

The simulation framework [33], validated in chapter 2, was used to simulate the 3D mass models into 2D and DBT patient images. The same system, described in section 2.2.1, was used in this study. The modifications applied to the simulation framework are discussed in the text and the flowchart in Fig. 4.7 describes the simulation framework applied to the mass simulation.

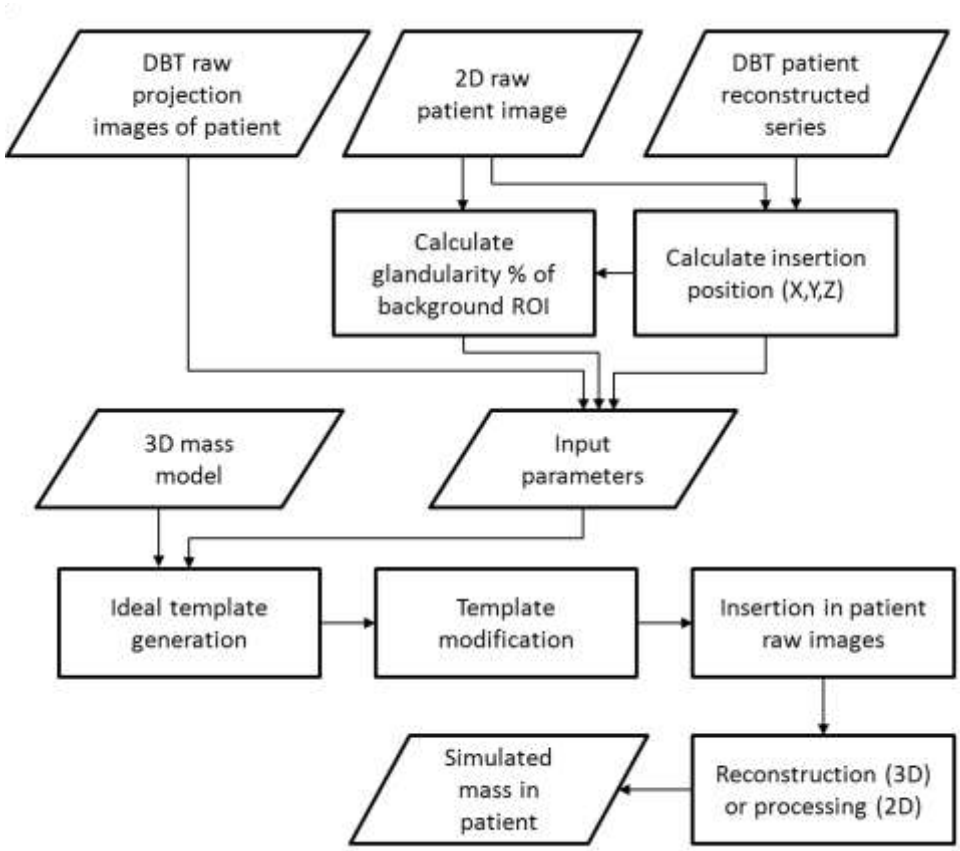


Fig. 4.7. Flow chart of the simulation framework to simulate mass into 2D and DBT patient images.

For the object simulation part, the input was the 3D mass model. Glandular tissue was used as substitute material for masses. The breast was assumed to be a mixture of glandular and adipose tissues with attenuation coefficients obtained from the work of Hammerstein *et al*. [89]. To estimate the background composition, a local implementation based upon the method derived from Kaufhold *et al.* [90] was used to estimate the glandularity percentage from the 2D raw patient image. Input parameters for the glandularity estimations are the mean pixel value in a region of interest (ROI) at the insertion position in the 2D projection, the tube voltage, the tube current exposure time, the anode/filter combination and the compressed breast thickness. The latter data were extracted from the DICOM header.

For the template modification part, The MTF curves were measured in the front-back and tube-travel directions in 2D acquisition mode, as explained in the previous chapters, and the MTF in the tube-travel direction was modified to include blurring from the tube motion in the left-right direction across the detector [35, 91] and to include the effect of the oblique incidence of the x-rays to the detector

surface, based on the model described by Que and Rowlands [53]. An approximate 2D MTF function for every projection angle was generated by combining the MTF curves in the tube travel direction and the front-back direction as described by Konstantinidis *et al.* [92].

In this study, the SPR values for the 2D mode (grid-in) and the DBT mode (grid-out) were taken from the work of Salvagnini *et al*. [57]. Further modifications, insertion and reconstructions/processing proceed as explained in the previous chapters.

4.2.4. Validation of the simulation

4.2.4.1. Image datasets

Image data from a total of 129 patients were used in the validation studies (2D FFDM and DBT) ; all patients had been imaged in 2D and tomosynthesis on the same system. Only raw format images of the mediolateral (MLO) views of tomosynthesis and 2D mammography were used. Seventy-four out of the 129 cases contained a total of 74 masses. Fifty-four out of the 74 masses formed the group of real masses as diagnosed from the radiological report and the other 20 cases were used for training purposes. Fifty five mass models were simulated into the images of the 55 patient cases without masses. All the mass models that had been simulated into 2D mammograms were then simulated into the DBT projections of the same patient in the same location allowing a paired analysis. The distribution of the masses used in the validation studies are described in Table 4.4.

Table 4.4. The distribution of the masses used in the validation studies and the training studies in terms of real and simulated population in the 2D and DBT studies.

			Real masses	**Simulated masses**	**Total**
Validation studies		2D	54	55	109
		DBT	54	44	98
Training sessions	1. Training to work with	2D	10	-	10
	DBT	DBT	10	-	10
	2. Training to work with	2D	10	-	10
	the software	DBT	10	-	10
Total		2D	74	55	129
(validation + training)		DBT	74	44	118

4.2.4.2. Observer study

Two observer studies were conducted independently, one for 2D digital mammography, the other for breast tomosynthesis. For each of these, radiologists were asked to score the realism of all simulated and real masses on a 5-point scale: 1. Mass is unrealistic, 2. Mass is slightly unrealistic, 3. No preference, 4. Mass is slightly realistic, 5. Mass is realistic. The radiologists were also asked to score the BIRADS of the mass from 1 to 5 (see Table 4.1) assuming that the mass is real in order not to bias their assessment.

Five radiologists participated in both studies of which three had more than ten years of experience in screening and diagnostic mammography (further assigned as readers A, B and C). The other two radiologists were final-year students under training (readers D and E). The three experienced radiologists had worked with breast tomosynthesis for more than four years and had participated in a previous tomosynthesis study [34]. Two training phases were conducted before the start of the validation studies: 1. training to work with DBT; 2. training to work with the software.

1. Training to work with DBT: ten patient cases were assigned to this training session (see Table 4.4). Only readers D and E assessed both the 2D and DBT images from the PACS. The goal of this session was to train the less experienced radiologists to read the tomosynthesis series, to be familiar with the appearance of the masses in the reconstructed images, to compare the lesions in DBT to 2D digital mammography and to gain confidence with this new modality. The radiologist was free to select/change the image magnification and the window width and level. They were also free to discuss these cases with the other radiologists.
2. Training to work with the software: a software package SARA [60] was used as a visualization tool in the observer studies. This training session was divided into two sessions, one for 2D and the other for the tomosynthesis study. Ten patient cases were included in these sessions containing 10 real masses. The aim of these sessions was to train all participating radiologists to the use of the software.

The validation studies were performed with a high resolution 5 megapixel monitor (Barco MDNG5121CB) in a routine mammography reading room. For the 2D study, the 109 real and simulated masses were divided into four reading sessions with the order of the masses randomly distributed within the sessions. The images were presented to the radiologists one by one having the whole breast (only the MLO view) shown with a rectangle that could be toggled on and off around the mass in question. The radiologists were free to zoom in/out, move within the mammogram and change the window/level settings without time limitations.

A similar subdivision of the 109 masses into four reading sessions was done for the tomosynthesis study. For easier visualization in the evaluation software, only 17 reconstructed planes around the central "in-focus" plane of the mass in question (+ 8 planes before and 8 planes after) were presented (1 mm plane spacing) per case. The radiologists were asked to evaluate the whole mass with the freedom to scroll within the 17 planes, zooming and changing window, level and width (Fig. 4.8). After radiologist A finished reading the DBT images, the results were used to filter out some "bad cases". A bad case was a simulated mass in DBT with realism score less than 4. Eleven cases were found and removed from the DBT study to end up with a total of 98 masses with 44 simulated masses and 54 real masses. Only these 98 cases were used in the validation study and were evaluated by the other radiologists.

The insertion positions of the simulated masses were chosen based on both the 2D processed image of the patients and the DBT reconstructed series. Every insertion position was selected by trial and error and accepted when the mass seemed to integrate well with the background in both 2D and DBT images. The distributions of the positions of real breast masses collected for the study for both the left and right breasts are shown in Fig. 4.9 along with the insertion positions of the simulated masses. Since each breast shape is different, the only conclusion from Fig. 4.9 is that the insertion positions of both real and simulated masses can be considered similar.

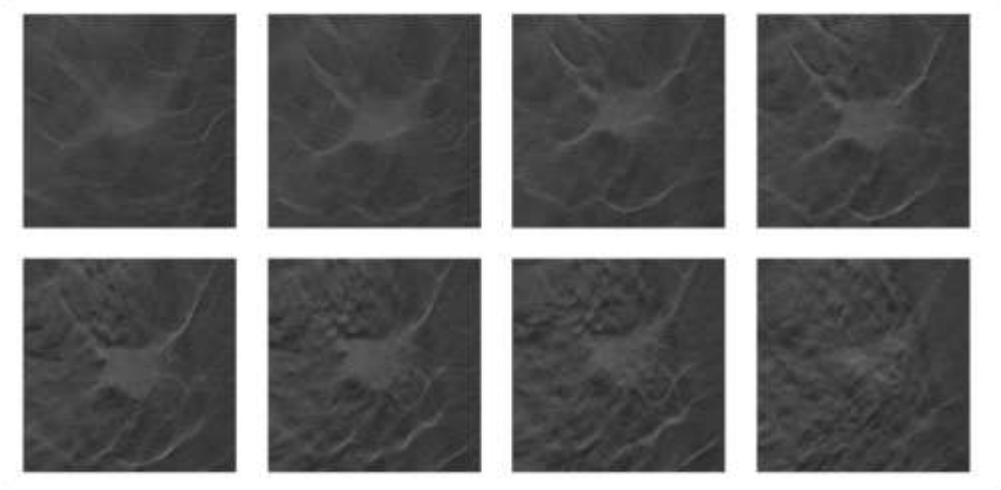


Fig. 4.8. Illustration of the different planes around the in-focus plane. These are 8 out of the 17 planes shown to the radiologists starting from the top left as 1, 3, 5…17.

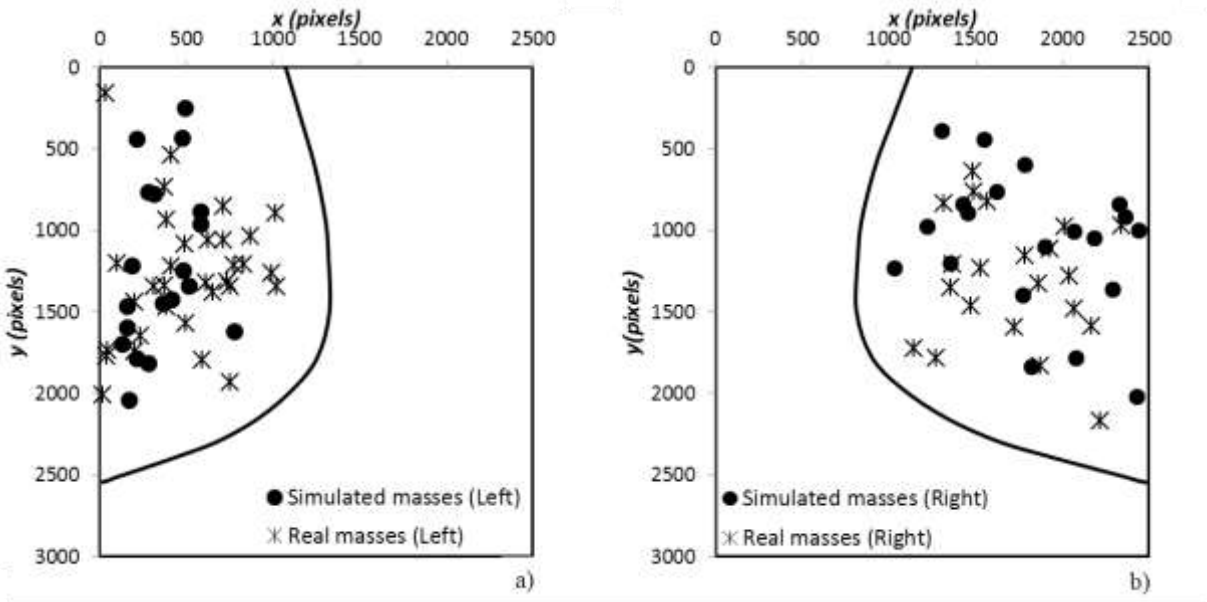


Fig. 4.9. The insertion positions in the 2D projection images of the patients for the real and simulated masses. a) the locations for left breasts, b) the locations for the right breasts.

The real masses selected for the validation studies had a wide diameter range of approximately 10 mm and therefore, we explored a variety of simulated diameter sizes within the same range (Table 4.5). It has to be noted that some spiculated masses were assigned a diameter zero because these were stellate lesions with no central core to be measured. The isotropic voxel sizes used to simulate spiculated and non-spiculated masses ranged between 0.03 to 0.05 mm depending on the required simulated diameter mass size as shown in Table 4.5.

Table 4.5. The diameters of the real and simulated masses used in the validation studies for spiculated and non-spiculated masses reported in terms of median values and inter-quartile range (IQR). The isotropic voxel sizes used to simulate spiculated and non-spiculated masses are also given.

		Median (mm)	**IQR (mm)**
Real masses	Non-spiculated	9.0	(8, 10)
	Spiculated	8.3	(7, 10.3)
Simulated masses	Non-spiculated	8.3	(8, 9)
	Spiculated	6.2	(5, 7)
Voxel sizes of the	Non-spiculated	0.040	(0.035, 0.04)
simulated masses	Spiculated	0.046	(0.044, 0.05)

4.2.4.3. Statistical analysis

Statistical analysis was performed with Receiver Operating Characteristic (ROC) [58] in order to check for differences in assessment between simulated and real masses. Similar to several studies in literature [34, 51, 59], we assume that the real and simulated lesions cannot be distinguished under the null hypothesis, the area under the ROC curve (AUC) should equal 0.5. The areas under the ROC curves were computed using the trapezoidal method along with the 95% confidence interval (CI) for each reader and for each modality using the function `perfcurve` from MATLAB (version 7.12, MathWorks, Natick, MA).

The inter-reader variability for the BIRADS scores was assessed among all observers for each modality using the Kendall's coefficient of concordance that ranges from 0 (no agreement) to 1 (complete agreement). This assessment of agreement was applied first to the simulated and real lesions separately, and then to the complete combined group of lesions for each modality.

For comparison of the BIRADS assessment between 2D and DBT, a paired comparison using the Wilcoxon signed rank test was separately performed for the real masses and for the simulated masses.

4.3. Results

4.3.1. Results of the 2D FFDM study

The results of the ROC analysis per observer are shown in Table 4.6 and Fig. 4.10. For all observers except observer C, no statistically significant difference was found between the realism scores of real and simulated lesions (all 95% CI included the value 0.5, $p>0.05$). Their AUCs varied from 0.54 to 0.66. The average of all readers' AUC was 0.58 (95% CI [0.50, 0.66], $p>0.05$). Since the 95% CI includes the value 0.5 ($p>0.05$), the null hypothesis cannot be rejected. Some examples of simulated and real cases in the 2D FFDM study are shown in Figs. 4.11 and 4.12. Figure 4.11 shows "well simulated" non-spiculated and spiculated masses respectively, i.e., a simulated mass interpreted as a real one with a mean realism score of more than 4.5 among the five radiologists. Figure 4.12 shows real masses that were interpreted as simulated with mean realism scores less than 3.5 among all radiologists.

Table 4.6. Area under the ROC curve (AUC) along with the 95% CI for the 2D FFDM study per observer; the last row gives the average AUC over all readers and corresponding 95% CI. An asterisk (*) indicates statistical significant difference between real and simulated masses.

Observer	Area under the ROC curve (AUC)	
	AUC	95% CI
A	0.54	[0.48, 0.62]
B	0.57	[0.50, 0.66]
C*	0.66	[0.55, 0.75]*
D	0.59	[0.49, 0.69]
E	0.56	[0.46, 0.65]
Average	0.58	[0.50, 0.66]

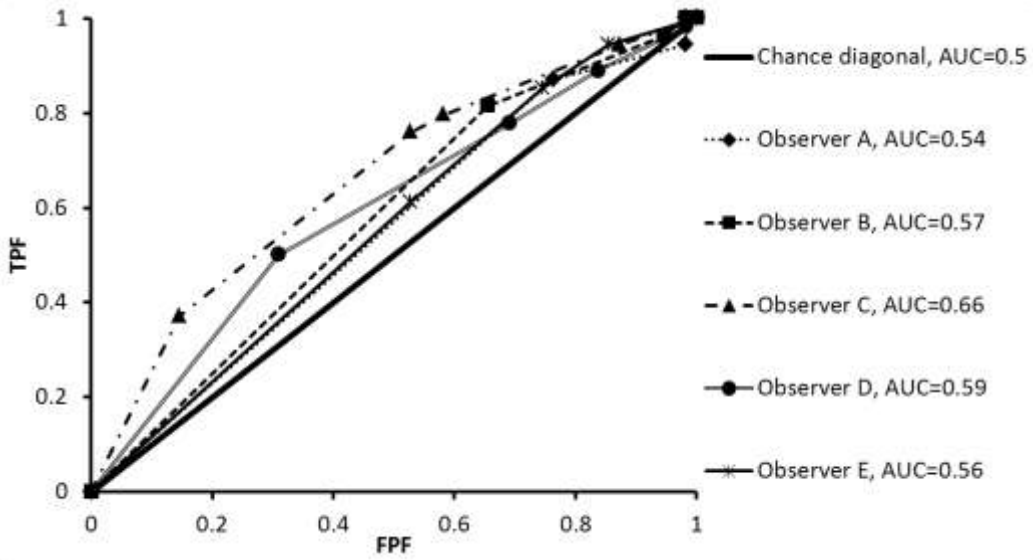


Fig. 4.10. ROC analysis of the 2D FFDM study with area under the ROC curve (AUC) per observer. The false positive fraction (FPF) is plotted against the true positive fraction (TPF).

The agreement of the BIRADS scores among the radiologists was assessed through the Kendall's coefficient which was 0.67 for all the masses, 0.62 for the simulated masses and 0.71 for the real masses, indicating good agreement between the radiologists.

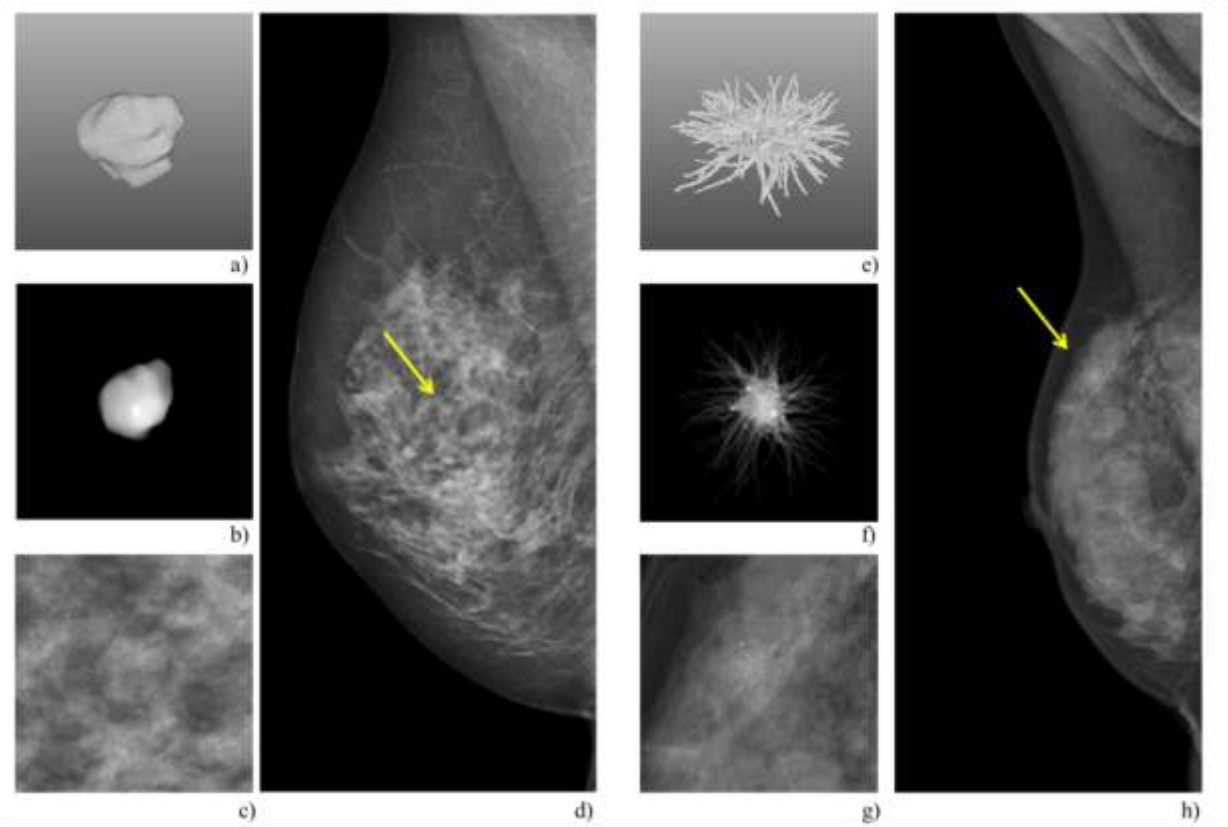


Fig. 4.11. Example of well simulated non-spiculated and spiculated mass models. The non-spiculated mass model had a mean realism score of 4.8 among all observers in the 2D study. a) The 3D non-spiculated mass model. b) The template (projection of mass in (a) in 2D before insertion). c) The mass in (b) after insertion in magnification view. d) The non-spiculated mass inserted in 2D with yellow arrow pointing to the location of insertion. The spiculated mass model had a mean realism score of 4.6 among all observers in the 2D study. e) The 3D spiculated mass model. f) The template (projection of (e) in 2D before insertion). f) The mass in (f) after insertion in magnification view. g) The spiculated mass inserted in 2D with yellow arrow pointing to the location of insertion.

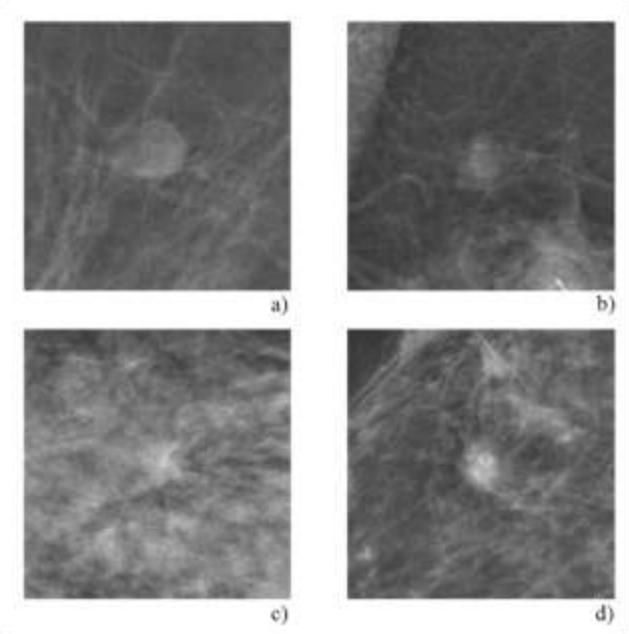


Fig. 4.12. Examples of low scored real masses with mean realism scores less than 3.5 among all observers in the 2D study.

4.3.2. Results of the tomosynthesis study

The results of the ROC analysis per observer are shown in Table 4.7 and Fig. 4.13. For three observers (A, C, D), statistically significant differences were found between the realism scores of real and simulated lesions (95% CI didn't include the value 0.5, $p<0.05$). Their AUCs varied from 0.60 to 0.77. The average of all readers' AUC was 0.67 (95% CI [0.55, 0.79], $p<0.05$) indicating significant differences. Some examples of simulated and real cases in the BT study are shown in Figs. 4.14 and 4.15. Figure 4.14 shows "well" simulated non-spiculated and spiculated masses respectively, i.e., simulated masses interpreted as real with a mean realism score more than 4.5 among the five radiologists. Figure 4.15 shows real masses that were interpreted as simulated with mean realism scores less than 3.5 among all radiologists.

The agreement of the BIRADS scores for the BT study among the radiologists was assessed through the Kendall's coefficient. This was 0.69 for all the masses, 0.57 for the simulated masses and 0.79 for the real masses, indicating good agreement between the radiologists.

Table 4.7. Area under the ROC curve (AUC) along with the 95% CI for the BT study per observer; last row reports the average AUC over all readers and corresponding 95% CI. An asterisk (*) indicates statistical significant difference between real and simulated masses.

Observer	Area under the ROC curve (AUC)	
	AUC	95% CI
A*	0.71	[0.60, 0.79]*
B	0.62	[0.50, 0.71]
C*	0.77	[0.66, 0.85]*
D*	0.65	[0.54, 0.74]*
E	0.60	[0.49, 0.70]
Average*	0.67	[0.55, 0.79]*

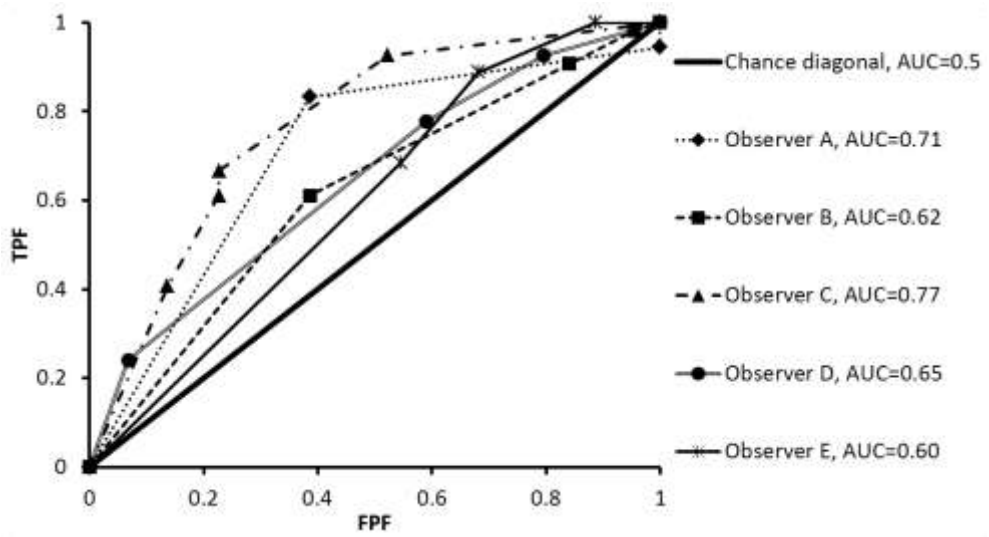


Fig. 4.13. ROC analysis of the tomosynthesis study with area under the ROC curve (AUC) per observer. The false positive fraction (FPF) is plotted against the true positive fraction (TPF).

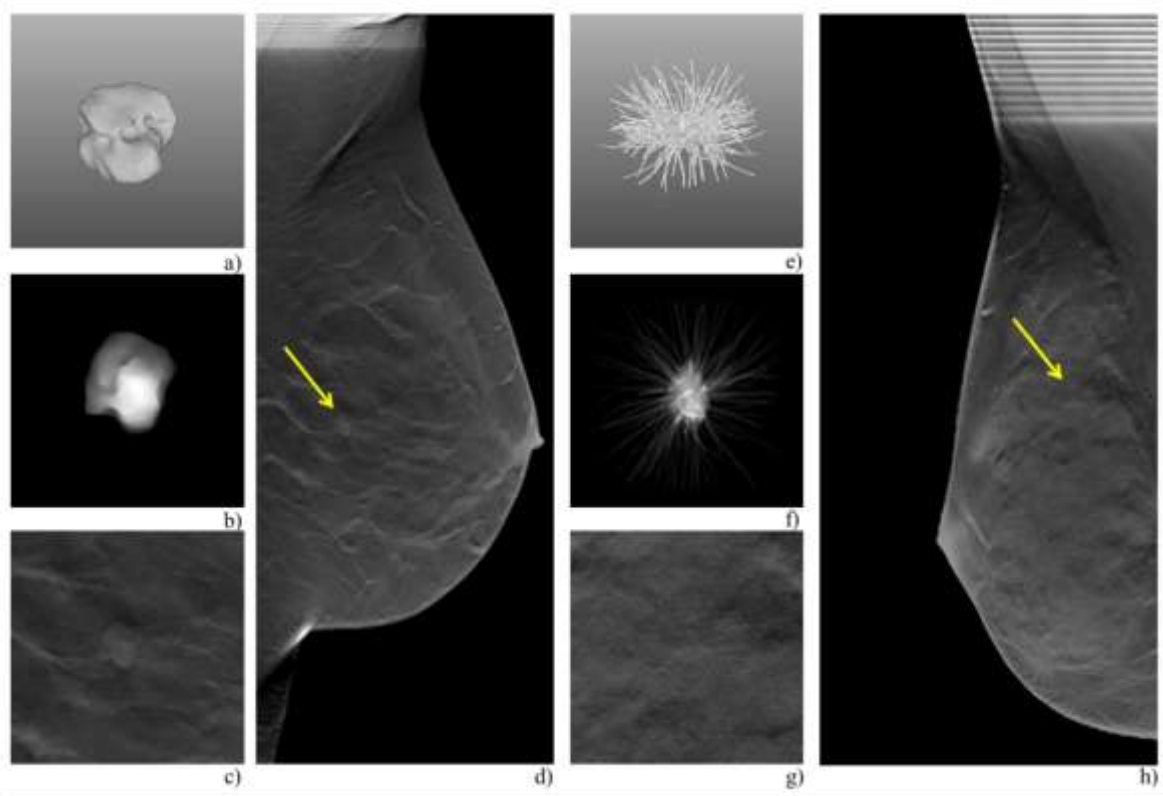


Fig. 4.14. Example of well simulated non-spiculated and spiculated mass models in the tomosynthesis study. The non-spiculated had a mean realism score of 4.8 among all observers. a) The 3D non-spiculated mass model. b) The template (central projection of (a) before insertion). c) The mass in (b) after insertion in magnification view. d) The non-spiculated mass in the in-focus plane in BT with yellow arrow pointing to the location of insertion. The spiculated mass model had a mean realism score of 5 among all observers in the BT study. e) The 3D spiculated mass model. f) The template (central projection of (e) before insertion). g) The mass in (f) after insertion in magnification view. h) The spiculated mass in the in-focus plane in BT with yellow arrow pointing to the location of insertion.

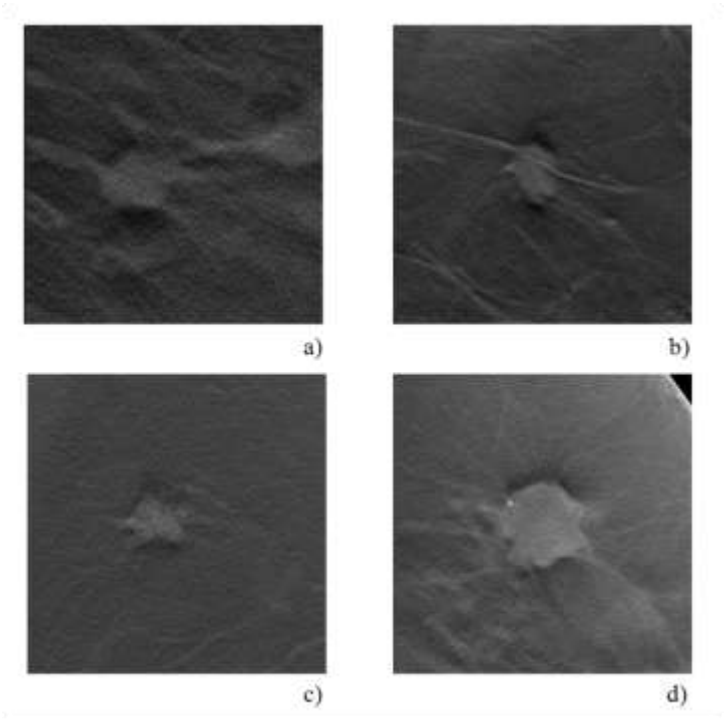


Fig. 4.15. Examples of low scored real masses shown in the in-focus reconstructed plane, with mean realism scores less than 3.5 among all observers in the tomosynthesis study.

4.3.3. Extra comparative results

Table 4.8 shows the scoring of the masses produced by pooling all observer results and using majority voting for each mass and for each modality separately. A mass was classified as "real" when the majority of the observers (3 or more) scored the realism of the lesion to be 4 or higher on the five point scale. The simulated masses' scores were 87% and 91% in 2D and DBT respectively compared to the scores of the real masses that were 96% and 94% in 2D and DBT. A mass was classified as "malignant" when the majority of the observers (3 or more) scored the BIRADS of the lesion to be 4 or higher on the five point scale. The results, based on the majority voting, of the 2D and DBT for the real and simulated masses, were all 48% indicating an almost even distribution for the masses used in the studies.

A paired analysis was applied to compare the BIRADS assessment of 2D and DBT for each observer for the 44 simulated masses and the 54 real masses. For a better understanding, the masses were classified as benign or malignant based on the majority voting of the DBT BIRADS scores, giving four paired analyses: 26 malignant real masses, 28 benign real masses, 21 malignant simulated masses and 23 benign simulated masses. The results are shown in Table 4.9. The table gives the mean and

standard deviation of the differences between the BIRADS scores in DBT and 2D for each observer and the corresponding Wilcoxon p-value. The mean values for the benign masses were mainly negative, indicating higher scores in 2D than in DBT and no statistically significant difference was found between their assessment of both the simulated and real benign masses. For the malignant masses, the mean values were positive, indicating higher scores in DBT than in 2D and three observers had statistically significant differences for their assessment of BIRADS in DBT compared to 2D for the real masses. Three observers also scored significantly differently for the simulated malignant masses with positive mean values. These results show the consistency of the BIRADS assessment of the simulated masses (both benign and malignant) compared to the real ones.

Table 4.8. Summary of pooling all observers' results using majority voting for each mass and for each modality (2D and BT). This shows the percentage of masses scored as "real" and the percentage of masses scored as "malignant" per mass group and per modality by the majority voting.

	Real masses		**Simulated masses**	
	2D	DBT	2D	DBT
No. of masses	54	54	55	44
Scored as real (%)	96%	94%	87%	91%
Scored as malignant (%)	48%	48%	48%	48%

Table 4.9. Paired analysis of 2D versus DBT in terms of the BIRADS scoring for the benign and malignant real and simulated masses. Reporting mean of the differences of the BIRADS scores with the standard deviations and the corresponding p-value obtained from applying the Wilcoxon signed rank test for each observer. An asterisk (*) indicates a statistically significant difference between the BIRADS scores in 2D and BT.

Observer	**Benign real masses (n=28)**		**Malignant real masses (n=26)**		**Benign simulated masses (n=23)**		**Malignant simulated masses (n=21)**	
	Mean±SD	p-value	Mean±SD	p-value	Mean±SD	p-value	Mean±SD	p-value
A	0.2 ± 1.1	0.347	0.7 ± 1.1	0.008*	0.3 ± 1.5	0.418	0.5 ± 1.0	0.048*
B	-0.2 ± 1.5	0.509	0.4 ± 1.2	0.114	-0.2 ± 1.9	0.653	0.9 ± 1.4	0.014*
C	0.0 ± 1.1	0.881	0.7 ± 1.5	0.027*	-0.5 ± 1.6	0.194	0.5 ± 1.2	0.078
D	-0.3 ± 1.2	0.308	0.8 ± 1.6	0.039*	-0.3 ± 1.5	0.374	0.1 ± 1.6	0.697
E	-0.3 ± 1.4	0.313	0.4 ± 1.3	0.118	0.0 ± 1.5	0.749	0.9 ± 1.4	0.017*

4.4. Discussion

The development and emergence of 3D modalities such as DBT and BCT is giving rise to research that investigates and optimizes their clinical performance especially when compared to the standard 2D modality. The simulation approach is a powerful tool to support such studies if all parts are validated. Along the same line, virtual trials as well as observer performance experiments are being organized for these technologies. Therefore, the development of 3D realistic and clinically-relevant mass models, with variety of morphological shapes and degrees of malignancies, is crucial. In this study, we introduced 3D non-spiculated and spiculated mass models. The realism of the appearance of the 3D mass models was successfully validated in both 2D and DBT. Since one of the aims to develop these masses was to conduct comparative observer studies, the validation studies were set up to allow a paired analysis: each mass model was simulated into 2D and DBT images of the same patient at the same position. The realism of the masses was evaluated using ROC and the AUCs with the 95% confidence interval were 0.58 (0.50, 0.66) for the 2D study and 0.67 (0.55, 0.79) for the DBT study. The simulation in DBT is more difficult than in 2D due to the 3D nature of the modality that enabled visualization of the margins of a mass better than in the 2D projected image and showed the spicules of the real masses to reach further compared to the typical lengths of spicules in our simulated ones. Moreover, the paired set-up required equal optimization of the appearance of a mass for both modalities which was not a straight forward task. In the partial simulation approach (simulating lesions in acquired patient images) the tissue at the location of the inserted lesion cannot be deleted. Therefore, some of the background structure may still be present. The use of our models in a full simulation approach is therefore even more promising.

In order to improve the simulations in DBT, we asked the experienced radiologists for their feedback. The main reasons for scoring a lesion to be unrealistic were the appearance of some masses with unrealistically higher density compared to background tissues, some masses looked superimposed, some masses with irregular or spiculated margins were obvious (indicated by a black halo) and did not integrate well with the background tissue. Fig. 4.16 shows examples of non-spiculated and spiculated masses scored as unrealistic in DBT while considered realistic in 2D, explaining the difficulties of optimizing the appearance of the masses in both modalities, the superimposed look and the black halo. The recommendations to improve the simulations in DBT include finding more realistic insertion positions to allow the lesion to integrate with the background and to use a pixel per pixel estimate of the background glandularity percentage that still does not exist in tomosynthesis projections.

Although the studies were carefully designed to cover different types of real and simulated masses, there are still limitations. 1) The insertion positions were selected and optimized by the first author after several trials and experience gained during the development of the masses. A better approach would have been to ask a non-participating radiologist in the studies to check the insertion positions and suggest better locations. Obviously, unrealistic positioning was corrected during the preparation phase though. 2) The 3D model matrix size was set to a maximum of 350x350x350 voxels to be usable in CPUs which prevented the spicule lengths to extend further as seen in some real masses in DBT. 3) The estimated glandularity for the background was based on the mean value from a ROI at the insertion site instead of a voxel based estimation.

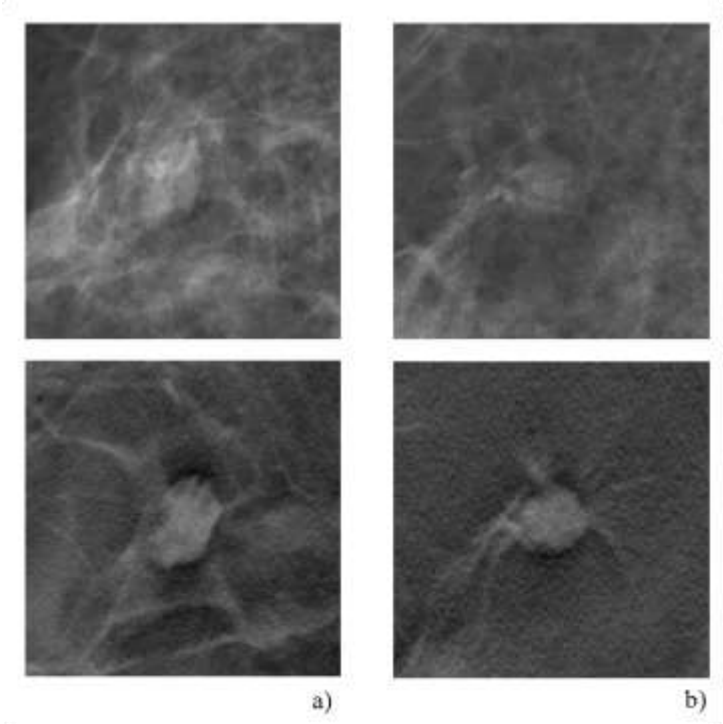


Fig. 4.16. Examples of simulated masses that were scored to be unrealistic in tomosynthesis and realistic in 2D, the top row are the masses in 2D and the bottom row are the masses inserted in BT showing the in-focus reconstructed plane. a) a non-spiculated mass with mean realism score among radiologists of 4.6 in 2D and 3 in BT. b) a spiculated mass with mean realism score of 4.4 in 2D and 3 in BT.

4.5. Conclusion

A database of 3D models of masses has been created and the realism of the appearance of the models has been validated; a ROC analysis was performed on 54 real and 55 simulated lesions in 2D mammograms. In DBT the validation used only a subset of the masses (44) compared to the 54 real masses. The results revealed high realism scores of the simulated masses in 2D and DBT. The area

under the ROC curves, averaged over all observers, was 0.54 (95% confidence interval [0.50, 0.66]) for the 2D study, and 0.67 (95% confidence interval [0.55, 0.79]) for the tomosynthesis study. The assessment of the BIRADS score for all observers indicated good agreement based on Kendall's coefficient that was 0.67 for the 2D scores and 0.69 for the DBT scores. The paired analysis of the BIRADS scores between 2D and DBT for each observer applied to the benign and malignant masses revealed consistent behavior for the real and simulated masses and confirmed the variety of shapes present in the studies. Following validation of the realism of the appearance of these models for the two modalities (2D FFDM and DBT) and after application of some improvements to the simulation in DBT, this database can now be used in future observer performance studies to optimize system parameters in both 2D FFDM digital mammography and breast tomosynthesis.

ACKNOWLEDGMENTS

This work is part of the OPTIMAM project which is funded by CR-UK & EPSRC Cancer Imaging Program in Surrey, in association with the MRC and Department of Health (England). We are grateful to the radiologists who participated in the studies: Dr. Andreas Van Steen, Dr. Sandra Postema, Dr. Liesbet Vermeiren, Dr. Anne-Sophie Vliegen. We would like to thank Dr. Thomas Mertelmeier from Siemens (Erlangen, Germany) for providing the reconstruction software "TomoEngine" and the reprocessing software "OpView2".

Chapter 4

Chapter 5

The influence of position within the breast on microcalcification detectability in continuous tube motion digital breast tomosynthesis

adapted from E. Shaheen et al., Proc. of SPIE Medical Imaging 2013

Chapter 5

Abstract

In digital breast tomosynthesis (DBT), the detectability and characterization of all lesions, especially microcalcifications, is still an issue under investigation. For DBT systems equipped with an x-ray tube that moves continuously during exposure, theory predicts some influence of the focal spot motion blur on detectability and diagnosis of small lesions, such as microcalcifications. Motion blur experienced by a lesion at some position in the breast is known to depend on the height of the lesion above the table within the breast. In this study, we investigated the influence of position above the table on microcalcification contrast and signal difference to noise ratio (SdNR) (as a surrogate for detectability) in tomosynthesis images, by means of a hybrid simulation method. Microcalcifications, represented by spheres of calcium with 400 μm diameter, were simulated into projection images of homogeneous objects and into anatomical backgrounds. The influence of system sharpness was included via the modulation transfer function (MTF) model that included detector, focus size, tube motion and x-ray oblique entry components. Results show contrast reductions for spheres at increasing heights above the detector in all datasets. For example, contrast drops of 31.5% and 43.1% for spheres inserted at 40 mm and 69 mm above the table compared to a sphere at 1 mm, respectively, were found for spheres simulated near the chest wall for homogeneous background. For the same cases, the corresponding drops in SdNR were 30.6% and 40.3%, respectively. Similar trends were also seen for sphere contrasts measured in anatomical backgrounds.

5.1. Introduction

Digital breast tomosynthesis (DBT) is a promising modality with possible applications in screening and diagnosis [22]. A major advantage of DBT over 2D planar mammography is the reduction of overlapping tissues in a given image plane resulting in improved lesion detectability [3]. A design choice for DBT systems is whether the tube moves while the x-ray tube is exposing, i.e. a flying focus versus a step and shoot design. In a previous study [5], it was shown that small lesions such as microcalcifications suffer from loss of contrast when the tube is moved continuously during exposure, compared to step and shoot mode. In that analysis, all lesions were simulated in the same position in the breast. The contrast of masses was not affected by tube motion [5].

In simulation approaches, whether full simulation [7, 29] or partial simulation [31, 34], many factors can be studied as long as they can be modelled in the simulation chain. Clinically realistic simulation in the frame of applications such as virtual clinical trials should include all aspects with eventual impact on performance parameters. Typically studied parameters are sharpness of the detector, dose

level and angular range. System sharpness as associated with flying focus systems is ignored in some simulation procedures [32].

This study extends our previous work that studied the effect of MTF for continuous tube motion versus step and shoot to include the influence of the position of small lesions in terms of height above the breast support table, and in different regions in the breast, with a focus on the simulation of small spheres in homogeneous and anatomical backgrounds acquired under continuous tube motion mode.

5.2. Materials and Methods

The hybrid simulation framework (chapter 2) [33], was used to simulate small spheres of calcium of diameter 400 μm into projection images of a homogeneous background (7 cm of PMMA) and anatomical backgrounds from real patient cases. For the anatomical backgrounds, acquisitions of three patients of 7 cm compressed breast thickness in craniocaudal (CC) view with different glandularity percentages estimated from an in-house program (as explained in chapter 4) were used: adipose breast of average 20% glandularity, medium with average 50% and dense breast with average of 70% glandularity. These percentages were used during the simulation procedure as explained later. Both phantom and patients were acquired with 31kV W/Rh.

Three datasets were simulated into different positions (x,y,z) of the background (phantom or patient) : 1. Chest wall side, 2. Nipple side and 3. Bottom right. These datasets were simulated at exactly the same x,y,z positions in both phantom and patient backgrounds. An extra set was simulated in the phantom: 4. Extreme nipple side to investigate an extreme region outside the breast borders, as shown in Fig. 5.1. Each dataset contained 8 simulated spheres inserted at different heights above the detector table: 1, 10, 20, 30, 40, 50, 60, 69 mm. The same system described in section 2.2.1. was used in this study.

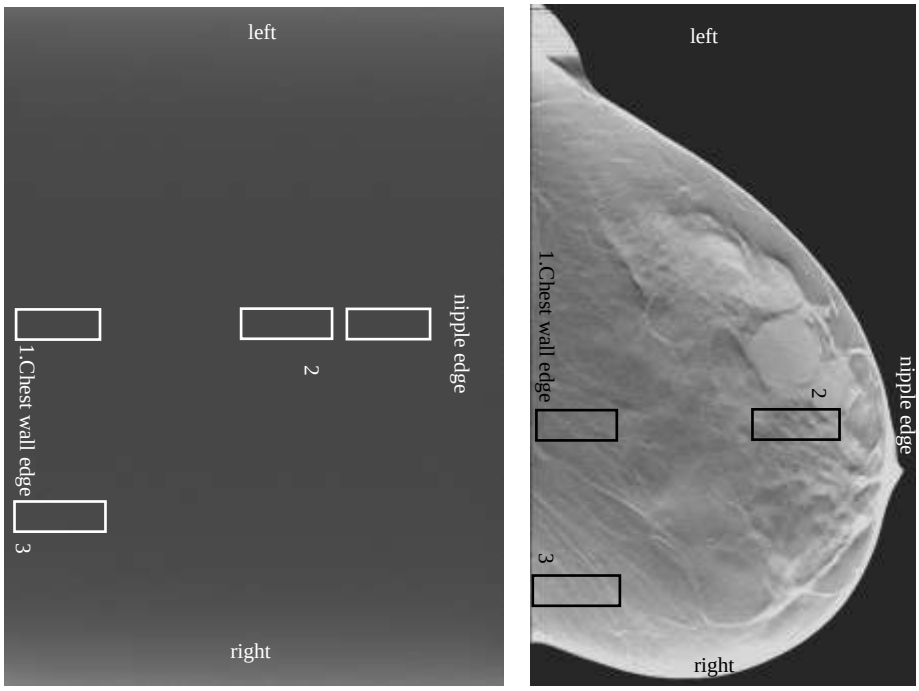


Fig. 5.1. Image of 7 cm PMMA(left), and patient (anatomical background) of 7cm (right) showing rectangles highlighting the regions representing the datasets simulated.

5.2.1. Simulation Framework

The simulation framework described in chapter 2 was used to simulate small objects into the 'FOR PROCESSING' projection images of DBT [33]. The simulation steps are the same as described in chapter 2, and the modifications applied in chapter 4. In this work, we improved the modelling of the sharpness of the inserted objects using a modified system MTF as discussed in the next section 5.2.2.

5.2.2. MTF modifications

In the initial simulation (chapter 2), the system MTF was measured with the edge at the table surface for the central projection image– this included the effect of tube motion for objects close to the table surface. A first adjustment (chapter 3) was a correction to the MTF to account for the oblique entry of the x-ray photons with respect to the *a*-Se detector in the tube-travel direction, based on the work of Que and Rowlands [53]. A recent study [91] has measured system MTF curves for DBT systems at different heights above the table and compared these data against the system MTF calculated from the detector MTF (measured at 0°, with a static tube and edge at the table) and MTF (sinc) functions that represent the focus size and focus motion. An example of a measured left-right detector MTF curve is

shown in Fig. 5.2; the figure also includes the correction for heights of 10 and 69 mm at 0° angle (i.e. the effect of oblique entry is not present).

Practically, the modifications applied to the MTF in this study were as follows: 1. Measure MTF in planar mammography mode (0° static tube, edge on the table) using the edge method for left-right and front-back directions. 2. Correct the front-back curve (chest wall to nipple direction) and left-right curve (tube-travel direction) using a sinc function for the extended focus size at the specified height. 3. Correct the left-right curve using a sinc function for focus motion at the specified height. 4. Apply the oblique entry correction to the left-right MTF curve. 5. Combine the modified front-back and left-right MTF curves to give a 2D MTF as described by Konstantinidis *et al.* [92].

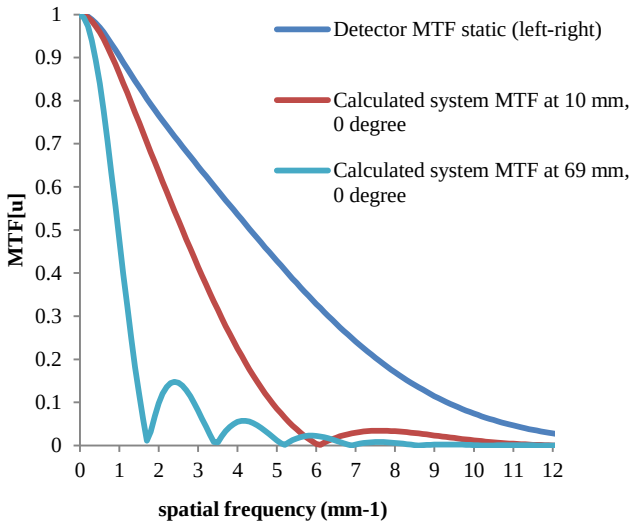


Fig. 5.2. The measured detector MTF of left right curve, corrected for heights at 10 mm and 69 mm for the zero degree angle.

5.2.3. Image analysis

Two figure of merits were used for analysis: the contrast (C) and the signal difference to noise ratio (SdNR). The contrast generated by the sphere was measured at the different stages of the modelling, to examine the influence of each parameter during the simulation. The first stage considered (1) was contrast in the ideal templates produced using simple ray tracing of the 3D Ca sphere model; the second stage (2) was template contrast after modification by the MTF (multiplication of the templates in the frequency domain by the height specific MTF). In step (3), the central projection image was modified to account for scattered radiation, and finally step (4) was the contrast in the in-focus plane

of the reconstructed images. Eq. 5.1 was used to estimate contrast in the templates (stages 1 and 2) while Eq. 5.2 was used for stages 3 and 4.

SdNR was measured in the in-focus plane of the reconstructed images (Eq. 5.3) as this metric is related to object detectability [93, 94].

$$C = 1 - \overline{PV}_{object} \tag{5.1}$$

$$C = \frac{|\overline{PV}_{object} - \overline{PV}_{background}|}{\overline{PV}_{background}} \tag{5.2}$$

$$SdNR = \frac{|\overline{PV}_{object} - \overline{PV}_{background}|}{\sigma_{background}} \tag{5.3}$$

where $\overline{PV}_{object}$ is the mean pixel value in the object, $\overline{PV}_{background}$ is the mean pixel value in the background region of interest (ROI), $\sigma_{background}$ is the standard deviation in the background ROI.

This model uses the MTF as a means of applying blurring from focus travel to small simulated objects – this will have a direct influence on contrast. SdNR is a function of contrast and noise; the noise itself is not affected by the blurring due to focus travel and hence contrast through the different planes experiences a blurring that depends on the object height above the table while noise remains unaffected.

5.3. Results

5.3.1. Results for homogeneous background

Figure 5.3 (left) shows the contrast measurements for three different stages: in the central projected ideal templates (original templates obtained from ray tracing the 3D model of the sphere), then after MTF corrections and finally in the central projection image for all spheres in the chest wall side dataset. This dataset was used as a reference position and to explain the effect of position in terms of height on contrast.

It is clear that the contrast is successively reduced at each stage. Figure 5.3 (right) presents all datasets in the PMMA and Table 5.1 shows the percentage drop in contrast between two stages: a) in going from original template to the inclusion of the MTF and b) in going from the MTF stage to the insertion stage with the inclusion of scatter, for spheres inserted at 1 mm, 40 mm, and 69 mm above

the table, respectively, for each dataset. For the MTF modification stage, contrast reduction ranged from 23.9% (for a sphere at 1mm) to a maximum of 51.3% (for a sphere at 69mm above the table). For the insertion stage, the contrast reduction ranged from 49.5% to 70.5%. It is clear that all datasets corresponding to all different positions are overlapping in Fig. 5.3 (right). This indicates that the position of the lesion in the breast has minor effects when compared to their heights.

Table 5.1. The drop in contrast due to MTF modification and due to scatter corrections (insertion into background) for all datasets and backgrounds averaged over dataset.

		Drop in C due to MTF (%)			**Drop in C due to scatter (%)**		
Background	**Dataset**	Min (sphere at 1mm height)	Median (sphere at 40 mm height)	Max (sphere at 69 mm height)	For sphere at 1mm height	For sphere at 40 mm height	For sphere at 69 mm height
PMMA	Chest wall side	23.9	47.3	51.2	49.5	54.8	55.0
	Nipple side	24.2	47.3	50.9	56.5	51.4	57.8
	Bottom right	24.2	47.5	51.3	66.5	60.6	70.5
	Extreme nipple side	24.4	47.4	51.2	52.3	54.8	56.9
Patient	Chest wall side	23.8	47.0	50.9	40.4	47.9	53.2
(adipose)	Nipple side	24.2	46.9	50.7	41.1	36.7	48.9
	Bottom right	24.4	47.4	51.5	48.8	47.5	56.6
Patient	Chest wall side	24.4	47.6	51.5	48.7	38.5	55.7
(medium)	Nipple side	24.5	47.4	51.4	59.9	57.1	61.0
	Bottom right	24.4	47.6	51.8	52.8	49.2	54.5
Patient	Chest wall side	24.6	47.4	51.4	60.0	50.0	70.0
(dense)	Nipple side	24.5	47.2	51.0	36.6	39.2	45.2
	Bottom right	24.2	47.4	51.7	34.5	38.5	58.7

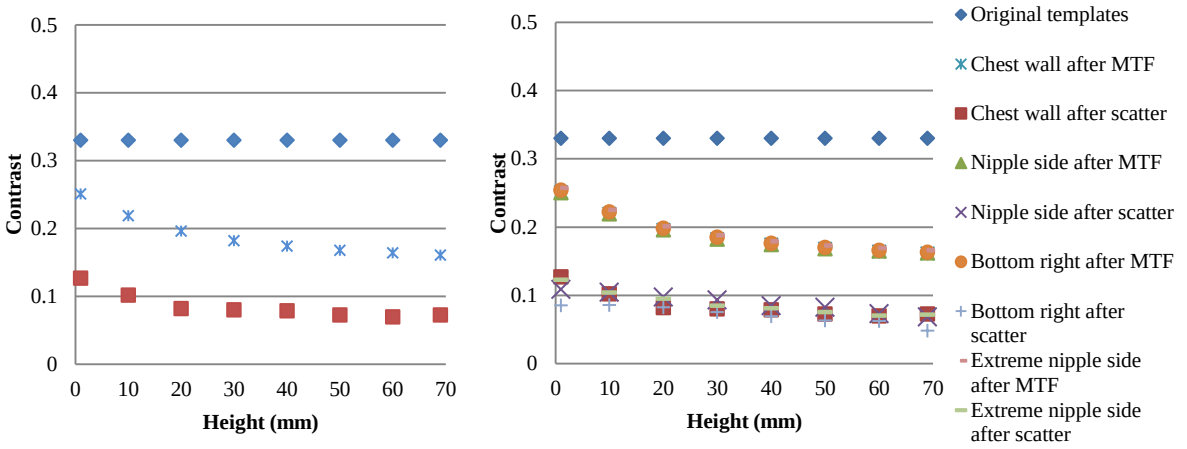


Fig. 5.3. Results of Contrast measurements for the Chest wall dataset (left) for PMMA for different stages: before MTF, after MTF and in the central projection image (after scatter) and for all the datasets (right).

Figure 5.4 shows the result of contrast and SdNR measurements in the in-focus reconstructed plane for all datasets. It is clear that the reduction in contrast with increasing height above the table translates into a reduction of SdNR as a function of height above the table. This SdNR measurement relates to detectability, especially as it is being used within one modality. Table 5.2 presents the contrast and the corresponding SdNR reductions for a sphere inserted at 1mm to a sphere inserted at 40mm, and for a sphere inserted at 1mm to a sphere inserted at 69mm, for all datasets.

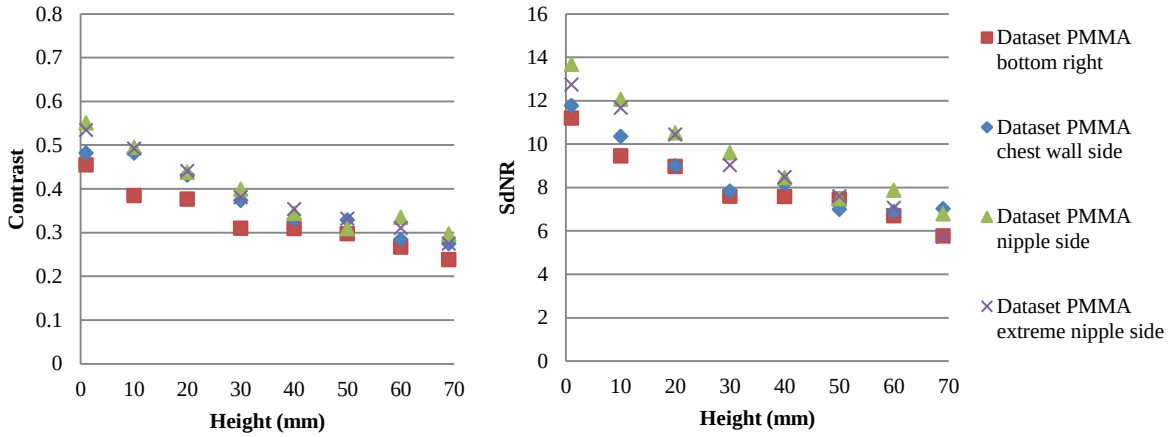


Fig. 5.4. Results of Contrast measurements (left) and SdNR (right) on all datasets for PMMA in the in-focus reconstructed plane.

Table 5.2. The drop in contrast and SdNR for a sphere inserted at 1mm compared to a sphere inserted at 40mm, and for a sphere inserted at 1mm compared to a sphere inserted at 69mm, for all PMMA datasets.

	Dataset	Drop in Contrast (%) for sphere at		Drop in SdNR (%) for sphere at	
		1mm→40mm	1mm→69mm	1mm→40mm	1mm→69mm
PMMA	Chest wall side	31.5	43.1	30.6	40.3
	Nipple side	37.5	46.1	38.2	50.4
	Bottom right	32.1	47.7	32.3	48.5
	Extreme nipple side	33.8	48.5	33.5	55.1

5.3.2. Results for anatomical background

The contrast measurements for the different stages (as explained in section 5.3.1) for all datasets in the three patients (adipose, medium and dense) are shown in Fig. 5.5 along with the contrast drop between every two consecutive stages in Table 5.1. For the MTF modification stage, contrast reduction ranged from 23.8% (for a sphere at 1mm) to a maximum of 51.8% (for a sphere at 69mm above the table). While for the insertion stage, the contrast drop ranged from 34.5% to 70%. It is clear that all datasets are overlapping in the stage of after MTF which indicates that the position of the lesion in the breast has minor effects.

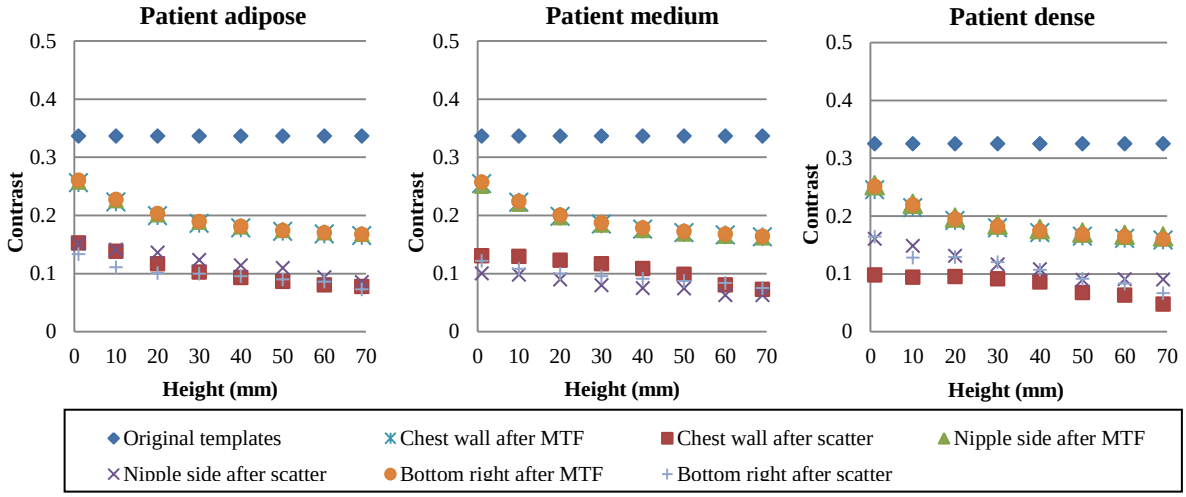


Fig. 5.5. Results of Contrast measurements for all datasets for all patients: adipose breast (20% glandularity), medium glandularity (50%), dense breast (70%glandularity) for different stages: before MTF, after MTF and in the central projection image (after scatter).

Figs. 5.6 and 5.7 show the results of contrast and SdNR, respectively, in the in-focus reconstructed plane in all datasets for all patient backgrounds. We expect the SdNR to exactly follow the contrast results, if the noise (standard deviation) remains constant in the different planes. Although there are obvious contrast reductions for spheres inserted at 69mm compared to a sphere inserted at 1mm (Table 5.3), for all datasets in all patient backgrounds, this contrast reduction was not directly reflected in the SdNR measurements in all patients sets (Fig. 5.7).

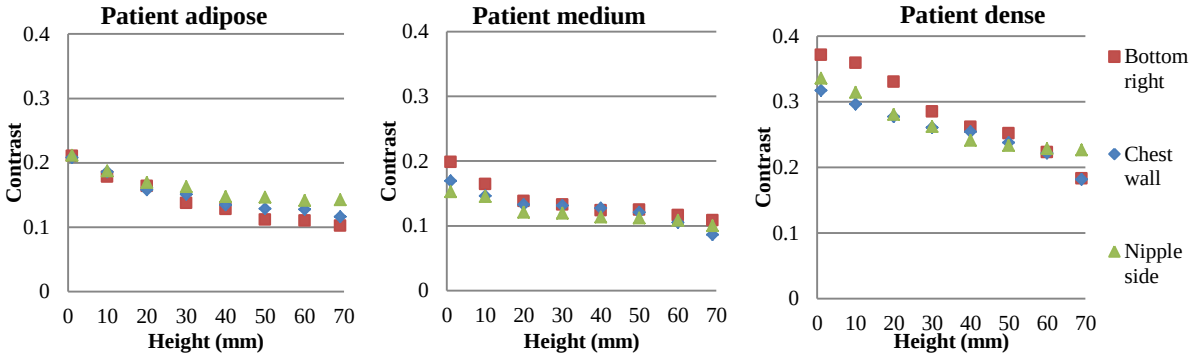


Fig. 5.6. Results of Contrast measurements in the in-focus reconstructed plane for all datasets for Patient adipose, medium and dense.

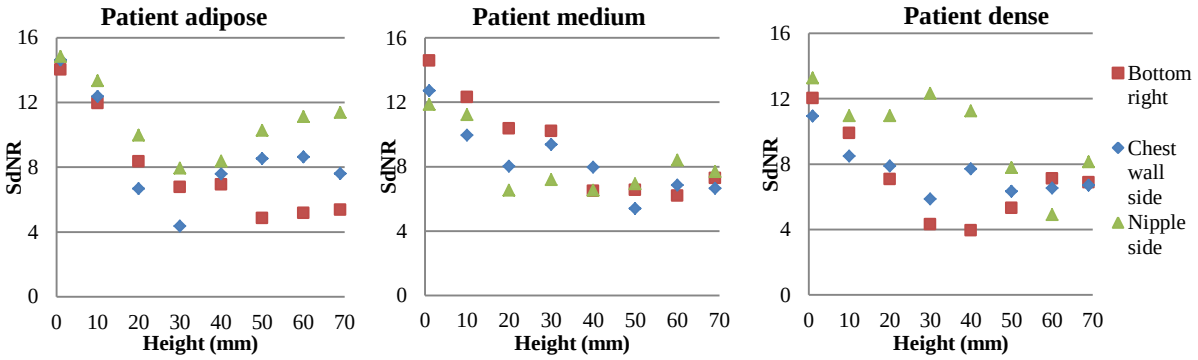


Fig. 5.7. Results of SdNR measurements in the in-focus reconstructed plane for all datasets for Patient adipose, medium and dense.

Table 5.3. The drop in contrast for a sphere inserted at 1mm compared to a sphere inserted at 40mm, and for a sphere inserted at 1mm compared to a sphere inserted at 69mm, for all Patients datasets.

Dataset	**Patient adipose**		**Patient medium**		**Patient dense**	
	Drop in Contrast (%) for sphere at		Drop in Contrast (%) for sphere at		Drop in Contrast (%) for sphere at	
	1→40mm	1→69mm	1→40mm	1→69mm	1→40mm	1→69mm
Chest wall side	35.4	44.2	25.0	49.1	19.7	42.8
Nipple side	30.2	32.7	25.7	34.3	28.2	32.5
Bottom right	39.3	51.5	37.5	45.4	29.6	50.8

This can be explained by the limitation of our background choices where it was not possible to find patient backgrounds with exactly the same glandularity distribution all over the breast. Moreover, SdNR measurements in patients are always difficult due to the impact of anatomical background. For the medium glandularity patient, SdNR reduction was clear but for the other two patients, the limitation applies. Therefore, we recommend the use of further detectability measurements such as model observers [11], and the application of the same methodology but in phantoms with different glandularity percentages. In this work, the choice of 400 μm diameter size was to ensure a consistent signal from which contrast could be measured consistently, however this limited the results to the larger diameter calcifications. Future work may examine the reduction of contrast and SdNR reduction for smaller calcification diameters, where we expect the contrast reduction to be greater. Comparisons of absolute contrast measured in the reconstructed planes between patients or patients and phantoms were not applicable due to differences in the final scaling applied to the reconstructed dataset. These in turn are a reflection of differences in the image content and dynamic range in the projection data for the different acquisition datasets.

5.4. Conclusion

This study has shown that a simulation that includes focus motion blur, via an MTF curve, demonstrates a notable reduction in microcalcification contrast for DBT systems with continuous tube motion during exposure. This has potential consequences for microcalcification detection and classification, an effect that increases with the increasing height of calcifications above the table surface. Systems with continuous motion should make a careful choice in terms of total irradiation time, angular range, tube travel speed and pulse duration. Results show contrast reductions in all

datasets. Contrast reductions of 31.5% and 43.1% were found for a sphere inserted at 1 mm height compared to an insertion, near the chest wall, at 40 mm and at 69 mm above the table, respectively, for homogeneous background. This drop was confirmed by SdNR measurements, where for the same example, the reductions were 30.6% and 40.3%, respectively. Similar values were also found for anatomical backgrounds for contrast measurements, but not directly translated in all anatomical backgrounds due to variations in the backgrounds.

ACKNOWLEDGEMENTS

This work is part of the OPTIMAM project and is supported by CR-UK & EPSRC Cancer Imaging Programme in Surrey, in association with the MRC and Department of Health (England).

Chapter 6

The morphology of microcalcifications in 2D digital mammography and breast tomosynthesis: is it different?

adapted from E. Shaheen et al., Proc. of IWDM, 362–368, (2012)

Chapter 6

Abstract

The development of new 3D imaging systems in the mammographic field is raising questions on its superiority of performance over 2D digital mammography in all aspects. Researchers are currently investigating the performance of digital breast tomosynthesis (DBT) compared to 2D digital mammography in terms of detectability of lesions (masses and microcalcifications) and diagnostic accuracy. Since a morphological description of the shape of microcalcifications is a determining factor for recalling the patient or not, we have investigated the efficiency of DBT in describing the morphology of microcalcifications within clusters compared to digital mammography. Four radiologists participated in the study and have described the shapes of microcalcifications in 71 clusters in 2D images and DBT series that were read in separate blinded sessions. An agreement test based on the kappa statistic was applied to evaluate the consistency of each reader's evaluation in 2D and DBT. An inter-rater variability test was also applied for each modality. Results have shown that there is good agreement between the observers' evaluations in these two modalities. The inter-rater test also revealed good agreement between the observers performance of assessment. In conclusion, this preliminary study has shown that the morphology of microcalcification clusters does not differ substantially in 2D versus DBT.

6.1. Introduction

The role of digital breast tomosynthesis (DBT) is not yet clear, whether DBT could be used in screening, or should be used for diagnostic evaluation only or in high risk women. Gennaro *et al.* [18] and Andersson *et al.* [14] have compared DBT to full-field digital mammography (FFDM) in terms of cancer detection and characterization. They have found that lesion conspicuity increases in DBT compared with FFDM and in Ref. [14] even concluded that DBT has a higher specificity for breast cancer detection. Moreover, Spangler *et al.* [23] have found that FFDM was more sensitive and specific than DBT for the detection of microcalcifications.

Microcalcifications are considered a characteristic sign for localization of malignancy [95]. Their morphology is one of the most important characteristics indicating malignancy or benignity [96]. Therefore, it is important for any mammographic system to allow not only detection but also correct categorization of breast microcalcifications since this impacts on the recall rate, an important indicator and performance measure for screening. Whereas most authors agree that DBT might have advantages in detecting and categorizing masses, its value for microcalcifications is not yet determined [14].

The aim of this study was to investigate the characterization of microcalcifications, by comparing the morphological description of microcalcifications within clusters between DBT and 2D FFDM in an observer performance study. Whereas most observer studies comparing these two modalities apply a side by side reading, where the observer evaluates each case with the presence of both 2D and tomosynthesis, in our study, we opted to evaluate both techniques in separate, blinded sessions.

6.2. Materials and Methods

The observer study was performed with a total of 71 microcalcification clusters (cases). Thirty five cases were patients with real microcalcification clusters and 36 were clusters simulated into patient images [34], as described in chapter 3. Simulated clusters have been used to increase the number of cases. The 3D microcalcification clusters were previously validated for their realistic appearance against real clusters in an observer performance study and the results showed no statistically significant difference between real and simulated clusters of microcalcifications in both 2D FFDM and DBT [34]. Each real microcalcification cluster present in a real breast was imaged in 2D and tomosynthesis using the same system (MAMMOMAT Inspiration TOMO, Siemens AG Healthcare, Erlangen, Germany). For the simulated clusters, the same patient background was also imaged using the same system and the cluster was simulated into the same position in 2D and DBT as described in chapter 3. All raw projection images were processed: the 2D images were processed using the default image processing software (OpView2, Siemens, Erlangen) and the projection tomosynthesis images were reconstructed using the Siemens software (TomoEngine, Siemens, Erlangen, Germany). The in-plane resolution of the reconstructed slices was 0.085 mm x 0.085 and the planes were reconstructed with 1 mm inter-slice distance. All patient doses in DBT are approximately double the dose of 2D FFDM, e.g. for a typical breast of thickness 45 mm the dose of one view 2D is 0.83 mGy and 1.79 mGy for one view DBT.

Four radiologists participated in the study. They evaluated the mediolateral (MLO) views of the patient cases using a user friendly software SARA^2 [60] on a high resolution 5 megapixel monitor (Barco MDNG5121CB) in our mammography reading room. The radiologists were trained first to read tomosynthesis series containing microcalcifications along with the 2D images from the PACS environment. Afterwards, they had two training sessions one for 2D and one for DBT, to learn the use of the SARA^2 software. In the subsequent observer study, they evaluated the 2D images in separate sessions from the DBT sessions to guarantee an independent reading. As the goal of this study was to evaluate the morphology of the microcalcifications within the clusters and not detectability, each

cluster was presented to the observers with a rectangle around it. The cases were presented to the radiologist one by one with the possibility to zoom in/out, change window level settings, toggle on/off the defining rectangle and scroll between planes in the case of DBT series. The radiologists were asked to describe the morphology of the microcalcifications in the cluster using one or more of the following four shapes: round, irregularly round, linear, and amorphous. Fig. 6.1 shows an example of the reading.

Comparison between the 2D and DBT setups was performed for each of the four shape categories (round, irregular round, linear and amorphous) separately using kappa statistics, and the observed agreement as well as the *p-value* were reported. No correction for multiple testing was performed in this preliminary study. Inter-rater agreement was assessed between each pair of the readers again using kappa statistics. A *p-value* of <0.05 was considered to indicate statistical significance.

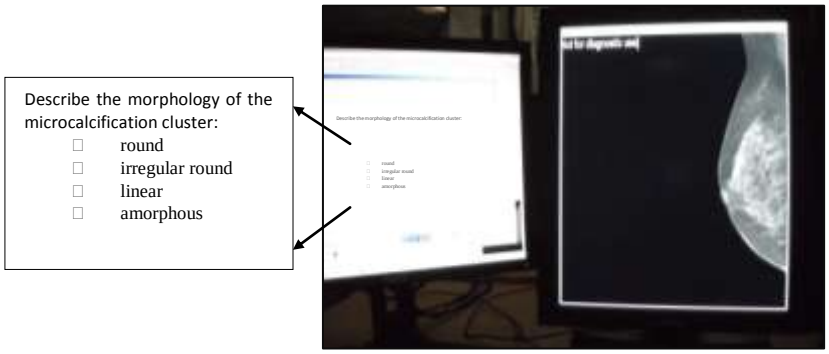


Fig. 6.1. Example of a reading session

6.3. Results

Table 6.1 shows the results of the agreement test based on kappa statistics with the corresponding p-value comparing 2D and DBT. The agreement was reported per shape category and observer. Agreements varied between 0.65 and 0.78 for the round shape, between 0.59 and 0.69 for the irregular round shape, between 0.75 and 0.83 for the linear shape, and between 0.76 and 1 for the amorphous shape. Generally, the agreement between the two modalities in the assessment of morphological description can be considered good, but only limited significance can be found in this limited population size.

Table 6.1. Agreement according to Kappa statistics and uncorrected p-values for comparison between 2D and DBT in terms of morphological description for the different shapes and observers

Observer	Shape							
	Round		Linear		Amorphous		Irregular round	
	Agreement	*p-value*	Agreement	*p-value*	Agreement	*p-value*	Agreement	*p-value*
1	0.65	0.028*	0.61	0.099	0.79	0.136	0.82	0.018*
2	0.79	0.070	0.62	0.094	0.83	0.679	0.79	0.310
3	0.78	0.478	0.59	0.315	0.75	0.690	1.00	<0.001*
4	0.78	0.025*	0.69	0.018*	0.80	0.119	0.76	0.209

Tables 6.2 and 6.3 show the results of inter-rater agreement between all observers for both 2D and DBT techniques, respectively. For 2D FFDM, the agreement varied from 0.73 to 0.78 indicating good agreement between readers in the evaluation of the morphological shapes of microcalcification clusters. For DBT, the agreement varied from 0.72 to 0.82 indicating good agreement as well.

Table 6.2. Inter-rater observed agreement based on kappa statistics between the four observers for all evaluations in 2D

2D FFDM				
Observer	**1**	**2**	**3**	**4**
1	1	0.78	0.76	0.76
2	-	1	0.76	0.73
3	-	-	1	0.76
4	-	-	-	1

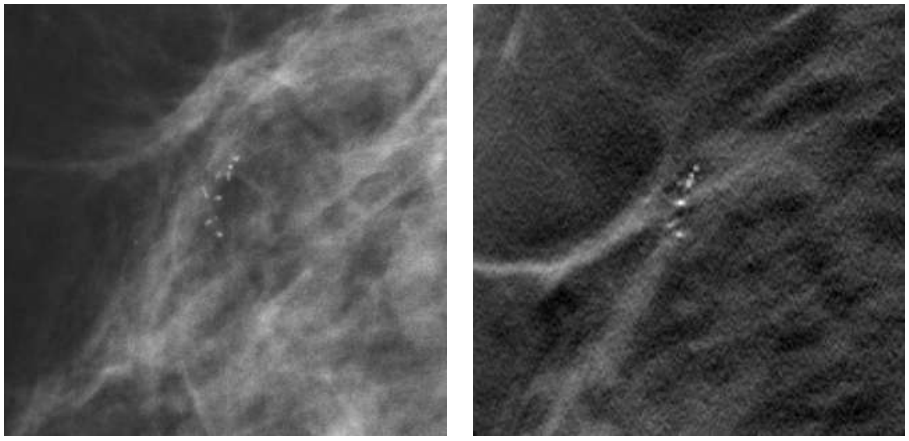
Table 6.3. Inter-rater observed agreement based on kappa statistics between the four observers for all evaluations in DBT

	DBT			
Observer	**1**	**2**	**3**	**4**
1	1	0.77	0.72	0.73
2	-	1	0.82	0.81
3	-	-	1	0.81
4	-	-	-	1

6.4. Discussion

The introduction of new technologies requires extensive research. Currently, a number of studies are investigating the role of DBT in the field of breast imaging. Some studies are interested in the physical development and optimization of the systems while other studies are evaluating the capabilities of DBT compared to 2D FFDM in clinical applications. Some studies show that DBT offers superior detectability for masses but is equal to 2D FFDM for microcalcifications. Other studies have focused on general detectability of microcalcifications and masses and found no significant difference between 2D and DBT [14, 18]. One major aspect that is not yet fully investigated is the characterization of microcalcifications. This study setup stems from the validation procedure of simulated microcalcification clusters when some differences were noticed between the appearance of clusters in 2D compared with the same clusters in DBT, as shown in Fig. 6.2. Fig. 6.2 a) is a microcalcification cluster in 2D and Fig. 6.2 b) is the same cluster in a DBT reconstructed plane (in focus plane). Table 6.4 shows the assessment of the cluster in Fig. 6.2 by all four observers in both 2D and DBT. It is clear that two observers have changed their interpretation from round to irregular round. The linear shape of calcifications was missed by all observers in DBT. When asked to provide feedback after having read all cases, the radiologists commented that generally the appearance of microcalcifications was slightly different in DBT when compared to 2D in terms of distribution over planes and reconstruction artifacts. But they didn't refuse the existence of reconstruction artifacts because they revealed the presence of irregularly shaped microcalcifications due to the peaks at the border of the microcalcification. The observers also commented that DBT images are noisier.

This study has focused on the morphological appearance of microcalcification clusters in 2D and DBT in highlighted regions. Although this was done intentionally, another approach would be to design the study as a search (detectability) and diagnostic task with a standardized descriptor such as Le Gal [97] or BIRADS [84]. Another limitation was the relatively low sample size, which was reflected in the non-significant correlation statistics (Table 6.1) where the *p-values* were reported for completeness. In future studies, a prospective setup in a larger patient population needs to be performed to further assess correlations between the 2D and DBT setups.



a) b)

Fig. 6.2**.** An example of a microcalcification cluster in a) 2D FFDM and in b) DBT reconstructed plane (in focus plane).

Table 6.4. The assessment of the cluster in Fig. 6.2 by all four observers in both 2D and DBT.

Observer	**Assessment in 2D FFDM**	**Assessment in DBT**
1	round	irregular round
2	round, irregular round, linear	round, irregular round
3	round	irregular round
4	round, linear	Round

6.5. Conclusion

A comparison in terms of morphological description was applied between 2D FFDM and DBT. A total of 71 microcalcification clusters were evaluated by four radiologists who were asked to describe the shapes of the microcalcifications in the clusters in terms of round, irregular round, linear or amorphous. The observer study was implemented in blinded sessions, where the observers read the 2D images in separate sessions from the DBT series. An agreement test was calculated using kappa statistics between the two modalities (2D and DBT) and the results showed good agreement among readers per shaped category indicating that the morphology of microcalcification clusters does not substantially differ in 2D versus DBT.

ACKNOWLEDGMENTS

This work is part of the OPTIMAM project which is funded by CR-UK & EPSRC Cancer Imaging Program in Surrey, in association with the MRC and Department of Health (England). The authors are grateful to the radiologists who participated in the study: Dr. Sandra Postema, Dr. Kirsten Joossens, and Dr. Riet D'Hauwe. The authors would like to thank Thomas Mertelmeier from Siemens (Erlangen, Germany) for providing the reconstruction software "TomoEngine" and the reprocessing software "OpView2."

Chapter 7

The evaluation of microcalcification clusters in 2D digital mammography and breast tomosynthesis in terms of morphology and number of calcifications in the cluster

adapted from E. Shaheen et al., *submitted to Brit. J. Radiol. (2014)*

Chapter 7

Abstract

Objectives: To evaluate microcalcification clusters in two-dimensional full field digital mammography (2D-FFDM) and digital breast tomosynthesis (DBT) in terms of morphology and number of calcifications detected in each cluster.

Methods: Six radiologists evaluated 51 microcalcification clusters in paired 2D-FFDM and DBT images. The two modalities were read in separate sessions to reduce bias. For each modality and for each cluster, the observers were asked to identify all the Le Gal types (morphology) present in the cluster and to count the number of visible microcalcifications. The McNemar test was applied, for each Le Gal type, to check for differences between the two modalities. For the evaluation of the number of calcifications in the clusters, the Wilcoxon signed rank test and linear regression were applied. Inter-reader variability were tested through the Kendall's coefficient of concordance.

Results: The morphology assessment between 2D and DBT, analysing each Le Gal type separately showed significantly different scores for only one observer for Le Gal types 2 and 5. All observers counted more calcifications in 2D-FFDM than DBT with the difference ranging between 31% and 47%. The Kendall's coefficients among the readers for the 2D was 0.473 and was 0.412 for DBT indicating moderate agreement.

Conclusions: A significant difference in the counted microcalcifications was found between 2D-FFDM and DBT whereas differences in morphological assessments were reader dependent. Therefore, the interpretation of microcalcifications in DBT should be subject to further investigation and training.

7.1. Introduction

Recently, Houssami and Skaane [98] reviewed the clinical studies conducted for DBT. Most of the studies were small scaled and they concluded that two-view DBT has at least equal or better accuracy than standard two-view 2D FFDM [12–14], and that the use of DBT as an additional technique to mammography in patients referred with an abnormal screening mammogram or with clinical symptoms increases accuracy [16, 17, 99]. In a population-based screening program by Skaane *et al.*[22], the use of mammography plus DBT resulted in a significantly higher cancer detection rate and an increased detection of more invasive cancers compared to 2D only. The question whether DBT can

replace standard mammography in screening population is still unanswered, requiring clinical studies evaluating the performance of DBT independently from 2D.

There are some concerns regarding the visualization of microcalcifications by DBT as found in the limited literature that specifically addresses the clinical assessment of microcalcifications by DBT [12, 14, 23]. The evaluation of microcalcifications in DBT is therefore a topic under investigation with no conclusive results so far.

Since one third of invasive tumours are diagnosed based on microcalcifications only [100] and the percentage of malignant calcifications varies from 10% to 40% based on biopsy [47], it is important that microcalcifications remain successfully detected and diagnosed in any new breast imaging modality. The characteristics that determine the categories of microcalcifications include: morphology, distribution, diameter, number of calcifications in the cluster, variability, etc. [47, 96]. Morphology is one of the most important characteristics indicative for the histologic origin of the calcifications [96]. Fondrinier *et al.* [47] studied the importance of the morphological aspect of microcalcifications in the assessment of breast lesions in a population of 211 microcalcification clusters with histological files**.** They found that morphology and number of microcalcifications are relevant for detecting malignancy.

With the need of clinical studies focusing on the assessment of microcalcification clusters in DBT independently from 2D, we evaluated 51 microcalcification clusters imaged in both 2D FFDM and DBT in terms of morphology and the number of calcifications visualized in the cluster. The study used a blinded set-up where the 2D images were read in separate sessions from the DBT series and followed by a paired analysis of the readings.

7.2. Materials and methods

7.2.1. Case selection

Image data were collected from our routinely acquired DBT cases. Approval of the ethics committee was obtained for DBT imaging. Informed consent was not required: the decision for DBT was left to the radiologist's clinical decision. Present study is retrospective and had not been planned when the cases were acquired. Within the period of August 2009 and July 2010, images were selected following specific criteria: each image should contain at least one cluster of microcalcifications, the patient should have been imaged by the same system in 2D and DBT, and the same view had to be available

in both modalities. Thirty six patients were found with a total of 51 microcalcification clusters (each cluster was considered a case). Each patient was imaged using the same Siemens Mammomat Inspiration system with DBT functionality (Siemens, Erlangen, Germany). For this system, the x-ray tube moves continuously over an angular range of ±25° and produces 25 projection images. Only the mediolateral oblique (MLO) view of tomosynthesis and the corresponding 2D mammography view were used in the study. Tomosynthesis projections were reconstructed using the Siemens reconstruction algorithm that is based on filtered back-projection (FBP) [45]. The in-plane pixel dimensions of the reconstructed slices were 0.085 x 0.085 mm and the planes were reconstructed with 1 mm inter-slice distance. The 2D images were processed by the default image processing algorithm for 2D digital mammography (OpView2, Siemens, Erlangen).

Mean glandular dose (MGD) in DBT was approximately two times higher than 2D FFDM, e.g. for a breast of thickness 45 mm, MGD for one view 2D FFDM was 0.83 mGy and 1.79 mGy for one view DBT.

7.2.2. Observer study

Six radiologists participated in the study, five had more than 10 years of experience in screening and diagnostic mammography; the sixth radiologist was a resident in her last year. Three out of the five experienced radiologists had worked with DBT for about two years. Reading was conducted independently for 2D digital mammography and for DBT. The study was performed with a high resolution 5 megapixel monitor (Barco MDNG5121CB, Barco NV, Belgium) in a routine mammography reading room using a software platform for observer performance studies ($SARA^2$) [60]. A training phase, divided into two parts, preceded the study. In the first part, the radiologists were trained by reading ten clinical patient cases containing clusters of microcalcifications (both 2D FFDM and DBT series) retrieved from the picture archiving and communication system (PACS) . The second part focused on the use of the $SARA^2$ software: each radiologist read ten clusters of microcalcifications, per modality, using the software and answered the study questions. None of the training cases were used in the further study.

As the aim of the study was to evaluate the morphological aspects of the clusters rather than detectability, all clusters were visible in both 2D FFDM and DBT, with a rectangle delineating cluster locations within the whole breast. The rectangle could be toggled on and off and served as a pointer to the cluster to avoid ambiguity when more than one cluster was present in the same image. For each case, the whole breast was shown in 2D and in DBT. For the DBT series, the radiologists were asked

to fully evaluate the whole cluster with the freedom to scroll within the planes, zooming or adapting window level and width. This set-up reflected the daily reading situation. For the review of the 2D images, the 51 cases were read in a one hour session. The radiologists were free to zoom, pan and change the window level. The review of the DBT images took place on average 12 days after the 2D sessions, and the 51 clusters were divided into two reading sessions (each one hour).

The radiologists were asked to answer questions describing the morphology and the number of microcalcifications. Since the pathology reports (assumed to be the gold standard) were not available for all clusters and do not correlate with the individual calcifications within the cluster, we have restricted the study to a pure comparison between the two modalities with 2D FFDM considered as the reference. The morphology was assessed using the Le Gal descriptor [97] which is widely used in France and Europe [47]. The Le Gal classification guides the identification of characteristically benign or malignant calcifications as shown in Ref. [47] and allows a detailed and clinically relevant description of the calcifications in terms of their morphology. The readers were familiar with this classification scheme. Fig. 7.1 shows the Le Gal classification: Le Gal type 1 is annular, type 2 is regularly punctiform, type 3 is too fine for précising the shape (dust), type 4 is irregularly punctiform and type 5 is vermicular (linear). Table 7.1 explains the malignancy percentage associated with every morphological shape as found in Refs. [47, 97].

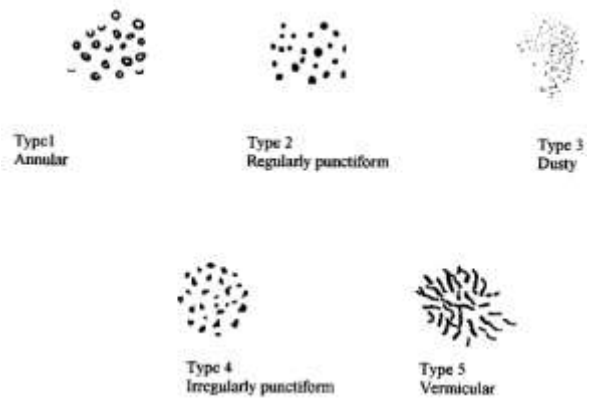


Fig. 7.1. Description of the morphology of microcalcifications following the Le Gal classification [97]

Table 7.1. The five Le Gal types and the corresponding malignancy percentage as reported by Le Gal et al. [97], and Frondinier et al. [47].

Le Gal type	**Malignancy (%) as found by Le Gal et al.**[97]	**Malignancy (%) as found by Frondinier et al.** [47]
(1) Annular	0	0
(2) Regularly punctiform	22	27
(3) Too fine for precising the shape (dusty)	40	32
(4) Irregularly punctiform	66	65
(5) Vermicular (linear)	100	100

In practice, both the morphology of microcalcifications and the number of microcalcifications in a cluster play an important role in the management decision whether it is follow-up, stereotactic biopsy or surgical biopsy [47]. The radiologists were therefore asked to answer two questions per cluster and per modality. It has to be noted that a cluster could contain more than one calcification shape (Le Gal type). The following questions were asked:

1. "Which Le Gal types are recognized in the cluster: Le Gal 1, Le Gal 2, Le Gal 3, Le Gal 4, Le Gal 5?" This was a multiple choice question and multiple answers were allowed. The question aimed to evaluate whether the microcalcifications were presented differently in both modalities and whether the modalities enabled a full morphological description of all microcalcifications in the cluster.
2. "How many calcifications can you see in the cluster?" The reader was asked to count the number of microcalcifications even if the microcalcifications were spread over different planes. When the number of microcalcifications exceeded 15, it was agreed to stop counting and put 15 as maximum number.

7.2.3. Statistical analysis

With the subjectivity of the morphological assessment, the paired McNemar test was applied to check for differences between 2D FFDM and DBT for each of the five Le Gal classes and per observer. The 2D FFDM was considered our ground truth and *p-value* of <0.05 was considered a significant difference of the assessment of Le Gal per observer between the two modalities.

For the second question, two tests were applied. A Wilcoxon signed-rank test was used to compare the number of microcalcifications per observer between 2D FFDM and DBT. A second analysis used linear regression and assumed that the gold standard was the 2D FFDM. The number of microcalcifications counted in the DBT cases was plotted against the number of calcifications in 2D FFDM. In this test, the R^2 was reported along with the slope of the fit in an attempt to find a relationship between the counts of microcalcifications in DBT compared to 2D FFDM.

The inter-reader variability for the Le Gal types was assessed for all observers using the Kendall's coefficient of concordance per modality.

7.3. Results

7.3.1. Results of the morphological description

Analysis of the results from the first question is shown in Table 7.2. The *p-values* from the McNemar test indicate no statistically significant difference between 2D FFDM and DBT for all observers for all Le Gal types except for two values: Le Gal type 2 for observer 1 and Le Gal type 5 for observer 6.

The results of the inter-reader observer variability using the Kendall's coefficient was 0.473 for 2D FFDM and 0.412 for DBT, indicating moderate agreement between the assessment of the observers for the Le Gal for both modalities.

Table 7.2. The p-values of the McNemar test applied per Le Gal type per observer between 2D FFDM and DBT.

Observer	**McNemar *p-values***				
	Le Gal 1	**Le Gal 2**	**Le Gal 3**	**Le Gal 4**	**Le Gal 5**
1	1.000	0.027*	0.289	0.581	0.726
2	0.250	0.383	0.064	1.000	0.227
3	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	0.210	0.688	0.250
5	0.059	0.690	0.134	0.070	0.500
6	1.000	0.210	0.343	1.000	0.002*

*indicates significant difference (p>0.05)

7.3.2. Results for the number of microcalcifications

The results of the two tests conducted for the second question are shown in Table 7.3. The output of the Wilcoxon test revealed a statistically significant difference between the counts of the microcalcifications in 2D FFDM when compared to DBT, for five out of six observers. A linear regression test was used to quantify the difference in counts. Based on the slope of the linear fit per observer, the number of microcalcifications counted in 2D FFDM was higher than in DBT for every observer, by between 31% and 47%.

Table 7.3. The Wilcoxon test applied to the counted microcalcifications in 2D FFDM and DBT and results of the linear regression test, R^2 and the percentage by which the calcifications counted in 2D FFDM exceeded those counted in DBT.

Observer	***p-value* from Wilcoxon test**	**R^2**	**Counts in 2D FFDM more than DBT by**
1	0.009*	0.88	47 %
2	0.011*	0.74	37 %
3	0.007*	0.69	40 %
4	0.002*	0.65	36 %
5	0.144	0.71	31 %
6	0.014*	0.90	32 %

* indicates significant difference

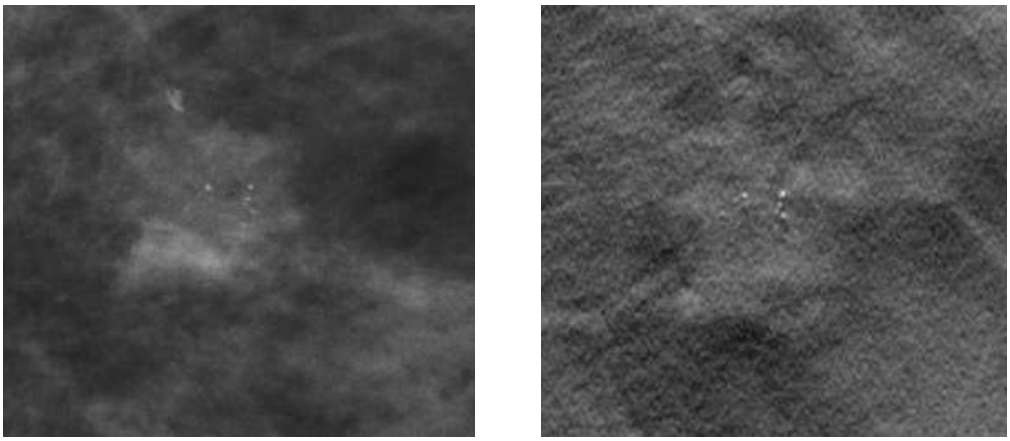
7.4. Discussion

Preliminary experience with the clinical use of DBT for the assessment of breast lesions in our hospital had revealed some issues with the imaging of microcalcifications. One observation was the differences of morphology between some calcifications in DBT and 2D FFDM. A second observation was that some calcifications seemed less visible in DBT: e.g. fine dust-like calcifications that were seen in 2D FFDM images were occasionally not seen in the reconstructed DBT planes. That's why we wanted to examine whether these issues were occasional, systematic or restricted to some limited cases.

In this study, the microcalcification clusters in 2D FFDM and DBT were evaluated in terms of morphology based on the Le Gal descriptor [97] and the number of calcifications in the cluster as these parameters are important for management follow-up decisions for the microcalcification clusters [47]. The study was designed for a paired set-up with DBT series read separately from 2D FFDM to

allow the evaluation of DBT as a stand-alone system. The comparison was applied between the two modalities tracking the differences in morphology and the changes in number of calcifications seen.

The results for the morphological evaluation for the individual analysis of every Le Gal class showed that only types 2 and 5 were scored significantly different by one out of the six observers. This result is considered subjective and a systematic change in the morphological assessment could not be concluded. The results of counting the number of microcalcifications showed a significant difference for most of the observers (five out of six observers), with all of them counting more microcalcifications in 2D FFDM when compared to DBT. An example is shown in Fig. 7.2 and Table 7.4. According to the radiologists involved in the study, the DBT images contained reconstruction artefacts and were sometimes noisy; an example is shown in Fig. 7.2.



a) b)

Fig. 7.2. A microcalcification cluster in a) 2D FFDM scored by an observer as Le Gal types 2 and 3 and b) In-focus DBT plane scored by the same observer as Le Gal type 2

Table 7.4. The counts of the microcalcifications in the cluster shown in Fig. 7.2 in 2D FFDM versus DBT for all observers.

Observer	Counts in 2D FFDM	Counts in DBT
1	10	5
2	10	4
3	8	5
4	10	5
5	9	5
6	10	4

Possible factors that may have an impact on microcalcification visualisation in DBT systems are the blurring due to the continuous motion of the tube during exposure [35, 91], scatter due to the lack of an anti-scatter grid [101] and the FBP reconstruction algorithm. Some practical solutions are recommended including the use of post-acquisition scatter reduction algorithms as proposed by Sechopoulos *et al.* [44] or the use of a grid may help to reduce the influence of scatter. The use of shorter x-ray pulses, ultimately step and shoot, could minimize the focal spot motion blur [91]. Another solution is the optimization of reconstruction algorithms for specific tasks such as microcalcifications [11].

The present study showed the non-inferiority of DBT when compared to 2D FFDM for the evaluation of morphology. Overall, evaluating and diagnosing microcalcifications with DBT should at present be done with great care and the radiological community could benefit from specific training in this aspect. Our results do not contradict the results found by Kopans *et al* [102] because they used a different DBT system with a tube that moves in a step and shoot mode, their set-up was not blinded (2D images were reviewed side by side with the DBT images) and they did not fully describe morphology or number of calcifications counted as in our study.

This work was an attempt to examine whether there are issues with microcalcification visualization in DBT. The focus of the study was to have a detailed description of the shapes and numbers of calcifications in a purely descriptive analysis. That was our main reason for using the Le Gal classification system rather than Breast Imaging-Reporting and Data System (BI-RADS) [84]. As a next step, a study could be performed to check detectability and diagnosis using a larger dataset and applying the BI-RADS classification along with Le Gal.

Some limitations of the study include: the visualization of the clusters in planes with a separation other than 1mm was not tested. Only one DBT system was investigated. There was a lack of pathology reports to be used as gold standard. Results of the present study cannot be extrapolated to systems with very different design parameters. Although the agreement for both modalities between the observers scoring the Le Gal type was moderate, higher agreement would be appreciated. This can be explained by the fact that the scoring is a subjective task. This confounding factor is not new for breast imaging studies. A previous study assessing microcalcifications using Le Gal classification in screening mammograms had shown fair agreement between two observers [103]. In the study by Crewson [104] on reader agreement studies, it was found that the agreement test values on the same

set of elements may vary a lot from one reader pair to another even when the observed agreement is relatively stable.

7.5. Conclusions

This paper evaluated microcalcification clusters in DBT compared to 2D FFDM. Present results are indicative for a potential clinical impact of describing clusters of microcalcifications in DBT versus 2D. The morphology assessment showed the non-inferiority of DBT compared to 2D FFDM when describing the different Le Gal types. A significant difference between these imaging techniques was found when the radiologists were asked to count the microcalcifications in the clusters, with significantly more calcifications seen in 2D FFDM. While these are results based on a small-scale study, this work highlights some issues of DBT regarding the visualisation of microcalcifications that should be further investigated.

ACKNOWLEDGMENTS

This work is part of the OPTIMAM project which is funded by Cancer Research-UK & EPSRC Cancer Imaging Program in Surrey, in association with the MRC and Department of Health (England). The authors are grateful to the radiologists who participated in the study: Dr. Andreas Van Steen, Dr. Sandra Postema, Dr. Kirsten Joossens, Dr. Ilse Vervloessem and Dr. Riet D'Hauwe.

Chapter 7

Chapter 8

Discussion and future perspectives

This doctoral work presented a new methodology to investigate and optimize the performance of the newly introduced 3D modality digital breast tomosynthesis (DBT) by setting-up and validating three novel simulation modules: the simulation framework, the 3D microcalcification clusters and the 3D mass models. The objective was to develop and validate accurate, efficient and flexible simulation tools. With the objective fulfilled, as explained in chapters 2 to 4, each module could be used either as a stand-alone module or as part of the task to simulate lesions into 2D and DBT images. We have presented the first applications that were implemented already within the scope of this PhD project but the power of the framework is not limited to what has been studied so far.

Simulation Framework

Optimizing the parameters influencing the image quality in DBT is still in progress. The developed simulation framework was validated by means of comparing the contrast of simulated inserts of phantoms to acquired projections of these inserts: no statistically significant difference was found between real and simulated objects. This framework is ready to be used in several applications to optimize the performance of DBT. One application was presented in chapter 5 where we investigated the influence of the detectability of microcalcifications when simulated at different heights from the detector table in both homogenous and anatomic backgrounds. The results showed that the contrast decreases when the height increases due to the blur of the focal spot motion. In addition, the effect of scatter and noise caused an overall extra drop in contrast. The results of this work pointed to possible clinical issues for the visibility and detection of microcalcifications in DBT systems equipped with a tube that moves continuously. We recommended further developments to overcome these issues.

Another specific application for the simulation framework is to guide phantom designs to test DBT systems by predicting the detectability of selected inserts. Different choices of materials of these inserts could be compared with the framework, next to their positions within any physically developed phantom. This framework was further used to simulate 3D lesions (microcalcification clusters and masses) into 2D and DBT images.

A user friendly simulation tool was developed using MATLAB. The tool is simple, fast and can be easily adapted to different 2D and DBT systems. Throughout the doctoral work, the simulation framework was optimized several times to achieve more accurate simulation through inclusion of the tube motion blur taking into consideration the height from the detector, the simulation of a polychromatic spectrum, a more realistic scatter estimation and background composition estimation of the breast. We could simply say that nothing is perfect, and there are some limitations to this

framework including: 1. This is a partial simulation approach that is not yet expanded to simulate breast phantoms, therefore, some parameters could not be investigated with this tool. 2. The tool is meant to simulate small 3D objects into acquired background, noise was so far not added as it is assumed that subtle objects don't affect the noise. However, extra noise could be added. 3. The insertion regions are limited to 80% of the breast area to avoid the borders of the breast where the scatter is no longer uniform. This framework can be extended to a full simulation approach with some modifications to include more detector properties (cascaded model), scatter maps, noise, other systems, other modalities (breast CT)…

Microcalcification clusters

A database of 44 3D models of microcalcification clusters was presented and validated for the realism of their appearance when simulated into 2D and DBT patient images in two observer studies. The models were segmented from reconstructed micro-CT images of biopsy specimens containing microcalcification clusters allowing the models to be used either as stand-alone or simulated into patient images by means of the simulation framework.

Our stand-alone 3D models were used in a joint study, as described by Alquran *et al.* (H. Alquran, E. Shaheen, et al. "Enhancement of 3D modeling and classification of microcalcification's clusters in breast computed tomography (BCT)", Proc. of SPIE, Vol. 9034 (2014)), where they were classified using a Support Vector Machine (SVM) classifier based on different features including compactness, size, radon transforms… The single microcalcifications were classified into different classes of Le Gal types [97] based on their morphological shapes. The next step for this work is to validate the classifier with microcalcification clusters acquired with breast computer tomography (BCT) system for use as computer aided diagnosis (CADx) algorithm. The 3D models can also be used as abnormalities in software breast phantom to conduct virtual clinical trials to optimize several parameters. The models were asked by a research group to be part of a breast phantom [105].

Since the 3D models were validated for realistic appearance, they can be used in observer performance studies with radiologists as observers. A number of applications to optimize and investigate the clinical performance of either 2D or DBT systems can be conducted such as comparing the detectability of lesions in 2D vs DBT, comparing different reconstruction algorithms, comparing different image processing algorithms, comparing different viewing conditions… The models are part of a joint on-going observer performance study to compare different 2D image processing techniques similar to Warren's et al. [106] earlier study. In their previous study, the authors had used 2D models

of microcalcification clusters [107] and real masses. The use of real masses was judged as problematic. In their current study, the radiologists asked to evaluate the cases using two-view images which required the simulation of 3D models of lesions (our microcalcification clusters and masses) for the estimation of the insertion positions in both views. Another large scale 2D clinical study is on-going in our group to compare the default automatic exposure control (AEC) settings with new AEC settings as proposed by Salvagnini et al. [108]. The aim of that PhD project is to reach constant detectability of lesions in different subgroups of breasts. This study will be using the simulation framework to simulate both our validated lesions: the microcalcification clusters and the masses.

A study was presented, chapter 6, that compared the morphology of microcalcification clusters in 2D versus DBT by means of real and simulated clusters. The results indicated that the morphology of microcalcification clusters does not substantially differ in 2D versus DBT. Some concerns were raised while presenting this work regarding the use of simulated clusters to evaluate morphology. Therefore, we conducted another study in 2D versus DBT to compare the morphological aspects of real microcalcification clusters only in terms of Le Gal descriptor and the number of calcifications within the clusters. The results revealed no systematic change in the morphological assessments, but a significant difference in the number of calcifications, with more calcifications seen in 2D. This work, in combination with the results of the study presented in chapter 5, highlight some issues of DBT regarding the visualisation of microcalcifications that should be further investigated. Recommendations to improve the visualization of calcifications in DBT include the use of post-acquisition scatter reduction algorithms as proposed by Sechopoulos et al. [44] or the use of a grid may help to reduce the influence of scatter. The use of shorter x-ray pulses, ultimately step and shoot, could minimize the focal spot motion blur [91]. Another solution is the optimization of reconstruction algorithms for specific tasks such as microcalcifications [11]. Detector MTF, focal spot motion and noise properties should be studied in a combined study that could be realized with our platform.

Breast masses

As the objective of this thesis was to optimize and investigate the clinical performance of DBT, it is not possible to rely only on one type of lesion. Therefore, we developed 3D models of breast masses. In order to include a variety of shapes, we proposed methods to simulate spiculated and non-spiculated masses. A database of 25 non-spiculated and 30 spiculated lesions was built and validated for the realism of the appearance of the models when simulated into 2D and DBT patient images as described in chapter 4. The results showed that the procedure to obtain realistic representations of the

masses is more challenging in DBT when compared to 2D due to the 3D nature of the modality. We found a number of factors independent from the design of the models that had an influence on the judgment of the appearance of the masses in DBT. One factor was the experience of the participating radiologists with the masses in DBT. Another aspect was the insertion positions and the integration of the mass with the background. Therefore, we were motivated to investigate some of these factors in a new pilot study in an attempt to improve the simulations and provide guidelines to future users (E. Shaheen, F. Bemelmans et al. "The investigation of different factors to optimize the simulation of 3D mass models in Breast Tomosynthesis", Proc. of IWDM 2014).

The 3D models could be used as part of virtual clinical trials by inclusion in software phantoms as explained earlier with the microcalcification clusters. All the potential applications presented in the section of the microcalcification clusters are the same, such as comparative studies of image processing in 2D or different reconstruction algorithms, or 2D vs DBT… In fact, all the previously mentioned observer studies to test the performance of a mammographic system can't be completed with only one type of lesions, it is a must that both microcalcification clusters and masses are present in the study to reflect the true clinical performance. The mass models are part of the two on-going 2D studies (the image processing comparison and the new AEC settings) as mentioned in the previous section.

Chapter 8

Bibliography

[1] Sechopoulos I (2013) A review of breast tomosynthesis. Part I. The image acquisition process. Med. Phys. 40(1):0143011–12

[2] Ziedses des Plantes BG (1932) New method of differentiation in radiography (planigraphy. Acta Radiol. 13:182–191

[3] Dobbins III JT (2009) Tomosynthesis imaging: At a translational crossroads. Med. Phys. 36(6):1956–1967

[4] Niklason T, Christian BT, Niklason LE, Kopans B, Opsahl-ong BH, Landberg E, Giardino A, Fitzgerald PF, Fobare F, et al. (1997) Digital Tomosynthesis in Breast Imaging. Radiology 205:399–406

[5] Shaheen E, Marshall N, Bosmans H Investigation of the effect of tube motion in breast tomosynthesis: continuous or step and shoot?, in: Proc. SPIE Vol. 7961, 2011: pp. 79611E1–79611E9

[6] Nishikawa RM, Reiser I, Seifi P, Vyborny CJ A new approach to digital breast tomosynthesis for breast cancer screening, in: Proc. SPIE Vol. 6510, 2007: p. 65103C–65108C

[7] Das M, Gifford HC, Connor JMO, Glick SJ (2009) Evaluation of a variable dose acquisition technique for microcalcification and mass detection in digital breast tomosynthesis. Med. Phys. 36(6):1976–1984

[8] Sechopoulos I (2013) A review of breast tomosynthesis. Part II. Image reconstruction , processing and analysis , and advanced applications. Med. Phys. 40(1):0143021–17

[9] Baker JA, Lo JY (2011) Breast Tomosynthesis: State-of-the-Art and Review of the Literature. Acad. Radiol. 18(10):1298–1310

[10] Zhang Y, Chan H, Sahiner B, Wei J, Goodsitt MM, Hadjiiski LM, Ge J, Zhou C (2006) A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis. Med. Phys. 33(10):3781–3795

[11] Michielsen K, Van Slambrouck K, Jerebko A, Nuyts J (2013) Patchwork reconstruction with resolution modeling for digital breast tomosynthesis. Med. Phys. 40(3):0311051–10

[12] Poplack SP, Tosteson TD, Kogel CA, Nagy HM (2007) Digital Breast Tomosynthesis: Initial Experience in 98 Women with Abnormal Digital Screening Mammography. Am. J. Roentgenol. 189(September):616–623

[13] Good WF, Abrams GS, Catullo VJ, Chough DM, Ganott MA, Hakim CM, Gur D (2008) Digital Breast Tomosynthesis: A Pilot Observer Study. Am. J. Roentgenol. (April):865–869

[14] Andersson I, Ikeda DM, Zackrisson S, Ruschin M, Svahn T, Timberg P, Tingberg A (2008) Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings. Eur. Radiol. 18:2817–2825

[15] Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, Rathfon GY, Sumkin JH, Zuley ML, et al. (2009) Digital Breast Tomosynthesis: Observer Performance Study. Am. J. Roentgenol. 193(August):586–591

[16] Teertstra HJ, Loo CE, Bosch MAAJ Van Den, Muller SH, Gilhuijs KGA (2010) Breast tomosynthesis in clinical practice: initial results. Eur. Radiol. 20:16–24

[17] Michell MJ, Iqbal A, Wasan RK, Evans DR, Peacock C, Lawinski CP, Douiri A, Wilson R, Whelehan P (2012) A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis. Clin. Radiol. 67(10):976–981

[18] Gennaro G, Toledano A, Maggio C di, Baldan E, Bezzon E, Grassa M La, Pescarini L, Polico I, Proietti A, et al. (2010) Digital breast tomosynthesis versus digital mammography: a clinical performance study. Eur. Radiol. 20:1545–1553

[19] Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M (2012) Two-View and Single-View Tomosynthesis versus Full-Field Digital Mammography: High- Resolution X-Ray Imaging Observer Study. Radiology 262(3):788–796

[20] Svahn TM, Chakraborty DP, Mattsson S, Andersson I, Ikeda D, Zackrisson S (2012) Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. Br. J. Radiol.

[21] Gur D, Zuley ML, Anello MI, Rathfon GY, Chough DM, Ganott MA, Hakim CM, Wallace L, Lu A, et al. (2012) Dose Reduction in Digital Breast Tomosynthesis (DBT) Screening using Synthetically Reconstructed Projection Images: An observer performance study. Acad. Radiol. 19(2):166–171

[22] Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, Izadi M, Jebsen IN, Jahr G, et al. (2013) Comparison of Digital Mammography alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program. Radiology 267(1):47–56

[23] Spangler ML, Zuley ML, Sumkin JH, Abrams G, Ganott MA, Hakim C, Perrin R, Chough DM, Shah R, et al. (2011) Detection and Classification of Calcifications on Digital Breast Tomosynthesis and 2D Digital Mammography: A Comparison. Am. J. Roentgenol. 196(February):320–324

[24] Humphrey LL, Helfand M, Chan BKS, Woolf SH (2002) Breast cancer screening: A summary of the evidence for the U.S. preventive services task force. Ann. Intern. Med. 137(5):347–367

[25] Ferlay J, Steliarova-foucher E, Lortet-tieulent J, Rosso S, Coebergh JWW, Comber H, Forman D, Bray F (2013) Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. Eur. J. Cancer 49:1374–1403

[26] (n.d.) NCI-funded Breast Cancer Surveillance Consortium co-operative agreement, "Performance Measures for 1,960,150 Screening Mammography Examinations from 2002 to 2006 by Age - based on BCSC data as of 2009,"; Available from: http://breastscreening.cancer.gov/d.

[27] Tingberg A, Zackrisson S (2011) Digital mammography and tomosynthesis for breast cancer diagnosis. Expert Opin. Med. Diagn. 5(6):517–526

[28] Young S, Bakic PR, Myers KJ, Jennings RJ, Park S (2013) A virtual trial framework for quantifying the detectability of masses in breast tomosynthesis projection data. Med. Phys. 40(5):0519141–05191415

[29] Gong X, Glick SJ, Liu B, Vedula AA, Thacker S (2006) A computer simulation study comparing lesion detection accuracy with digital mammography , breast tomosynthesis , and cone-beam CT breast imaging. Med. Phys. 33(4):1041–1052

[30] Zanca F, Jacobs J, Ongeval C Van, Claus F, Celis V, Geniets C, Provost V, Pauwels H, Marchal G, et al. (2009) Evaluation of clinical image processing algorithms used in digital mammography. Med. Phys. 36(3):765–775

[31] Timberg P, Bath M, Andersson I, Mattsson S, Tingberg A, Ruschin M (2010) In-plane visibility of lesions using breast tomosynthesis and digital mammography. Med. Phys. 37(11):5618–5626

[32] Timberg P, Bath M, Andersson I, Mattsson S, Tingberg A, Ruschin M (2012) Visibility of microcalcification clusters and masses in breast tomosynthesis image volumes and digital mammography: A 4AFC human observer study. Med. Phys. 39(5):2431–2437

[33] Shaheen E, Zanca F, Sisini F, Zhang G, Jacobs J, Bosmans H (2010) Simulation of 3D objects into breast tomosynthesis images. Radiat. Prot. Dosimetry 139(1-3):108–112

[34] Shaheen E, Van Ongeval C, Zanca F, Cockmartin L, Marshall N, Jacobs J, Young KC, Dance DR, Bosmans H (2011) The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis. Med. Phys. 38(12):6659–6671

[35] Shaheen E, Marshall NW, Bosmans H The influence of position within the breast on microcalcification detectability in continuous tube motion digital breast tomosynthesis, in: Proc. SPIE 8668, 2013: pp. 86684I1– 86684I10

[36] Shaheen E, Ongeval C Van, Zanca F, Cockmartin L, Marshall N, Keyzer F De, Young KC, Dance DR, Bosmans H The Morphology of Microcalcifications in 2D Digital Mammography and Breast Tomosynthesis: Is It Different?, in: Proc. IWDM, 2012: pp. 362–368

[37] Wu T, Stewart A, Stanton M, Mccauley T, Phillips W, Kopans DB, Moore RH, Eberhard JW, Opsahl-ong B, et al. (2003) Tomographic mammography using a limited number of low-dose cone- beam projection images. Med. Phys. 30(3):365–380

[38] Boone JM, Fewell TR, Jennings RJ (1997) Molybdenum , rhodium , and tungsten anode spectral models using interpolating polynomials with application to mammography. Med. Phys. 24(12):1863–1874

[39] M. J. Berger, J. H. Hubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar and DSZ (2005) XCOM: Photon Cross Sections Database. NIST Standard Reference Database 8.

[40] Nowotny R (1998) XMuDat: Photon attenuation data," computer program available at http://www-nds.iaea.org/reports/nds-195.htm.

[41] Siddon RL (1985) Fast calculation of the exact radiological path for a three-dimensional CT. Med. Phys. 12(2):252–255

[42] Samei E, Flynn MJ, Reimann DA (1998) A method for measuring the presampled MTF of digital radiographic systems using an edge test device. Med. Phys. 25(1):102–113

[43] Marshall NW (2006) A comparison between objective and subjective image quality measurements for a full field digital mammography system. Phys. Med. Biol. 51:2441–2463

[44] Sechopoulos I, Sankararaman Suryanarayanan SV, D'Orsi CJ, Karellas A (2007) Scatter radiation in digital tomosynthesis of the breast. Med. Phys. 34(2):564–576

[45] Mertelmeier T, Orman J, Haerer W, Dudam MK Optimizing filtered backprojection reconstruction for a breast tomosynthesis prototype device, in: SPIE, 2006: pp. 61420F1–61420F12

[46] Burnside ES, Ochsner JE, Fowler KJ, Fine JP, Salkowski LR, Rubin DL, Sisney GA (2007) Use of Microcalcification Descriptors in BI-RADS 4th Edition to Stratify Risk of Malignancy. Med. Phys. 242(2):388–395

[47] Fondrinier E, Lorimier G, Guerin-boblet V, Bertrand A, Mayras C, Dauver N (2002) Breast Microcalcifications: Multivariate Analysis of Radiologic and Clinical Factors for Carcinoma. World J. Surg. 26:290–296

[48] Zanca F, Chakraborty P, Ongeval C Van, Jacobs J, Claus F, Marchal G, Bosmans H (2008) An improved method for simulating microcalcifications in digital mammograms. Med. Phys. 35(9):4012–4018

[49] Ruschin M, Timberg P, Bath M, Hemdal B, Svahn T, Saunders RS, Samei E, Andersson I, Chakraborty DP, et al. (2007) Dose dependence of mass and microcalcification detection in digital mammography: Free response human observer studies. Med. Phys. 34(2):400–407

[50] Carton A, Bosmans H, Ongeval C Van, Souverijns G, Rogge F, Steen A Van, Marchal G (2003) Development and validation of a simulation procedure to study the visibility of micro calcifications in digital mammograms. Med. Phys. 30(8):2234–2240

[51] Saunders R, Samei E, Baker J, Delong D (2006) Simulation of Mammographic Lesions. Acad. Radiol. 13(7):860–870

[52] Fandos-Morera A, Prats-Esteve M, Tura-Soteras JM, Traveria-Cros A (1988) Breast Tumors: Composition of Microcalcifications. Radiology 169:325–327

[53] Que W, Rowlands JA (1995) X-ray imaging using amorphous selenium: Inherent spatial resolution. Med. Phys. 22(4):365–374

[54] Boone JM, Lindfors KK, CooperIII VN, Seibert JA (2000) Scatter/primary in mammography: Comprehensive results. Med. Phys. 27(10):2408–2416

[55] Hill-Kayser CE, Harris EER, Hwang W-T, Solin LJ (2006) Twenty-year incidence and patterns of contralateral breast cancer after breast conservation treatment with radiation. Int. J. Radiat. Oncol. Biol. Phys. 66(5):1313–1319

[56] Wu G, Mainprize JG, Boone JM, Yaffe MJ (2009) Evaluation of scatter effects on image quality for breast tomosynthesis Evaluation of scatter effects on image quality for breast tomosynthesis. Med. Phys. 36(10):4425–4432

[57] Salvagnini E, Bosmans H, Struelens L, Marshall NW (2012) Quantification of scattered radiation in projection mammography: Four practical methods compared. Med. Phys. 39(6):3167–3180

[58] Metz CE (2006) Receiver Operating Characteristic Analysis: A Tool for the Quantitative Evaluation of Observer Performance and Imaging Systems. J. Am. Coll. Radiol. 3(6):413–422

[59] Skiadopoulos S, Costaridou L, Kalogeropoulou CP, Likaki E, Livos L, Panayiotakis G (2003) Simulating the mammographic appearance of circumscribed lesions. Eur. Radiol. 13:1137–1147

[60] Jacobs J, Zanca F, Marchal G, Bosmans H Implementation of a novel software framework for increased efficiency in observer performance studies in digital radiology, in: Proc. RSNA, 2008

[61] Sechopoulos I, Ghetti C (2009) Optimization of the acquisition geometry in digital tomosynthesis of the breast. Med. Phys. 36(4):1199–1207

[62] Bliznakova K, Bliznakov Z, Bravou V, Kolitsi Z, Pallikarakis N (2003) A three-dimensional breast software phantom for mammography simulation. Phys. Med. Biol. 48:3699–3719

[63] Gang GJ, Tward D., Lee J, Siewerdsen JH (2010) Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. Med. Phys. 37(5):1948–1965

[64] Bakic PR, Albert M, Brzakovic D, Maidment ADA (2002) Mammogram synthesis using a 3D simulation . I . Breast tissue model and image acquisition simulation. Med. Phys. 29(9):2131–2139

[65] Bakic PR, Albert M, Brzakovic D, Maidment ADA (2002) Mammogram synthesis using a 3D simulation . II . Evaluation of synthetic mammogram texture. Med. Phys. 29(9):2140–2151

[66] Bakic PR, Albert M, Brzakovic D, Maidment ADA (2003) Mammogram synthesis using a three-dimensional simulation . III . Modeling and evaluation of the breast ductal network. Med. Phys. 30(7):1914–1925

[67] Bakic PR, Zhang C, Maidment ADA (2011) Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm. Med. Phys. 38(6):3165–3176

[68] Bliznakova K, Suryanarayanan S, Karellas A, Pallikarakis N (2010) Evaluation of an improved algorithm for producing realistic 3D breast software phantoms: Application for mammography. Med. Phys. 37(11):5604–5617

[69] Connor JMO, Das M, Dider CS, Mahd M, Glick SJ (2013) Generation of voxelized breast phantoms from surgical mastectomy specimens. Med. Phys. 40(4):0419151–0491512

[70] Li CM, Segars WP, Tourassi GD, Boone JM, Dobbins III JT (2009) Methodology for generating a 3D computerized breast phantom from empirical data. Med. Phys. 36(7):3122–3131

[71] Pokrajac DD, Maidment ADA, Bakic PR (2012) Optimized generation of high resolution breast anthropomorphic software phantoms. Med. Phys. 39(4):2290–2302

[72] Chen B, Shorey J, Saunders RS, Richard S, Thompson J, Nolte LW, Samei E (2011) An Anthropomorphic Breast Model for Breast Imaging Simulation and Optimization. Acad. Radiol. 18(5):536–546

[73] Hoeschen C, Fill U, Zankl M, Panzer W, Regulla D, Do W, Solutions M (2005) A HIGH-RESOLUTION VOXEL PHANTOM OF THE BREAST FOR DOSE CALCULATIONS IN MAMMOGRAPHY. Radiat. Prot. Dosimetry 114:406–409

[74] Reiser I, Nishikawa RM (2010) Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise. Med. Phys. 37(4):1591–1600

[75] Prionas ND, Huang S, Boone JM (2011) Experimentally determined spectral optimization for dedicated breast computed tomography. Med. Phys. 38(2):646–655

[76] Haka AS, Shafer-peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS (2002) Identifying Microcalcifications in Benign and Malignant Breast Lesions by Probing Differences in Their Chemical Composition Using Raman Spectroscopy. Cancer Res. 62:5375–5380

[77] Berks M, Barbosa D, Boggis C, Astley S Evaluating the realism of synthetically generated mammographic lesions: an observer study, in: Proc. SPIE Vol. 7627, 2010: pp. 7627041–76270411

[78] Berks M, Caulkin S, Rahim R, Boggis C, Astley S Statistical Appearance Models of Mammographic Masses, in: Proc. IWDM, 2008: pp. 401–408

[79] Ruschin M, Tingberg A, Bath M, Grahn A, Hakansson M, Hemdal B, Andersson I (2005) Using simple mathematical functions to simulate pathological structures-Input for digital mammography clinical trial. Radiat. Prot. Dosimetry 114(1-3):424–431

[80] Hintsala H, Bliznakova K, Pallikarakis N, Jämsä T Modelling of Irregular Breast Lesions, in: Proc. Int. Fed. Med. Electron. Biol. Eng. Vol. 25, 2009: pp. 2024–2027

[81] Ruschin M, Timberg P, Svahn T, Andersson I, Hemdal B, Mattsson S, Bath M, Tingberg A Improved in-plane visibility of tumors using breast tomosynthesis, in: Proc. SPIE Vol. 6510, 2007: pp. 65101J1–65101J11

[82] De Sisternes L, Zysk AM, Brankov JG, Wernick MN Development of a Computational Three-Dimensional Breast Lesion Phantom Model, in: Proc. SPIE Vol. 7622, 2010: pp. 7622051–7622058

[83] Rashidnasab A, Elangovan P, Yip M, Diaz O, Dance DR, Young KC, Wells K (2013) Simulation and assessment of realistic breast lesions using fractal growth models. Phys. Med. Biol. 58:5613–5627

[84] American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). 4th ed. Reston, VA: American College of Radiology, 2003

[85] Obenauer S, Hermann KP, Grabbe E (2005) Applications and literature review of the BI-RADS classification. Eur. Radiol. 15:1027–1036

[86] Murray CD (1926) The physiological principle of minimum work. I. The vascular system and the cost of blood volume. Proceeding Natl. Acad. Sci. United States Am. 12:207–214

[87] Rossitti S (1995) Energetic and spatial constraints of arterial networks. Arq. Neuro-Psiquiatr. 53(2):333–341

[88] Kurz H, Sandau K, Christ B (1997) On the bifurcation of blood vessels - Wilhelm Roux ' s Doctoral Thesis (Jena 1878) - A seminal work for biophysical modelling in developmental biology *. Ann. Anat. 179:33–36

[89] Hammerstein GR, Miller DW, White DR, Masterson ME, Woodard HQ, Laughlin JS (1979) Absorbed Radiation Dose in Mammography. Radiology 130:485–491

[90] Kaufhold J, Thomas JA, Eberhard JW, Galbo CE, Trotter DEG (2002) A calibration approach to glandular tissue composition estimation in digital mammography. Med. Phys. 29(8):1867–1880

[91] Marshall NW, Bosmans H (2012) Measurements of system sharpness for two digital breast tomosynthesis systems. Phys. Med. Biol. 57:7629–7650

[92] Konstantinidis AC, Olivo A, Speller RD (2011) Technical Note: Further development of a resolution modification routine for the simulation of the modulation transfer function of digital x-ray detectors. Med. Phys. 38(11):5916–5920

[93] Monnin P, Marshall NW, Bosmans H, Bochud FO, Verdun FR (2011) Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis. Phys. Med. Biol. 56:4221–4238

[94] Bernhardt P, Mertelmeier T, Hoheisel M (2006) X-ray spectrum optimization of full-field digital mammography: Simulation and phantom study. Med. Phys. 33(11):4337–4349

[95] Baker R, Rogers KD, Shepherd N, Stone N (2010) New relationships between breast microcalcifications and cancer. Br. J. Cancer 103(7):1034–1039

[96] Picca DA, Paredes ES de (2003) Calcifications in the breast: A radiologic perspective. Appl. Radiol. September:29–37

[97] Le Gal M, Chavanne G, Pellier D (1984) Valeur diagnostique des microcalcifications groupees decouvertes par mammographies. Bull. Cancer 71(1):57–64

[98] Houssami N, Skaane P (2013) Overview of the evidence on digital breast tomosynthesis in breast cancer detection. The Breast 22(2):101–108

[99] Bernardi D, Ciatto S, Pellegrini M, Anesi V, Burlon S, Cauli E, Depaoli M, Larentis L, Malesani V, et al. (2012) Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time. Br. J. Radiol. 85(December):1174–1178

[100] Weigel S, Decker T, Korsching E, Hungermann D, Böcker W, Heindel W (2010) Calcifications in Digital Mammographic Screening: Improvement of Early Detection of Invasive Breast Cancers? Radiology 255(3):738–745

[101] Wu G, Mainprize JG, Boone JM, Yaffe MJ (2009) Evaluation of scatter effects on image quality for breast tomosynthesis. Med. Phys. 36(10):4425–4432

[102] Kopans D, Gavenonis S, Halpern E, Moore R (2011) Calcifications in the Breast and Digital Breast Tomosynthesis. Breast J. 17(6):638–644

[103] Gulsum M, Demirkazik FB, Ariyurek M (2003) Evaluation of breast microcalcifications according to breast imaging reporting and data system criteria and Le Gal ' s classification. Eur. J. Radiol. 47:227–231

[104] Crewson PE (2005) Fundementals of Clinical Research for Radiologists: Reader agreement studies. Am. J. Roentgenol. 184(May):1391–1397

[105] Lau BA, Reiser I, Nishikawa RM (2012) A statistically defined anthropomorphic software breast phantom. Med. Phys. 39(6):3375–3385

[106] Warren LM, Given-Wilson RM, Wallis MG, Cooke J, Halling-Brown M, Mackenzie A, Chakraborty DP, Bosmans H, Dance DR, et al. (2014) Effect of image processing version on detection of non-calcification cancers in 2D digital mammography imaging. Am. J. Roentgenol. (accepted)

[107] Warren LM, Green FH, Shrestha L, Mackenzie A, Dance DR, Young KC (2013) Validation of simulation of calcifications for observer. Phys. Med. Biol. 58:N217–N228

[108] Salvagnini E, Bosmans H, Monnin P, Struelens L, Verdun F, Marshall NW The use of detectability indices as a means of Automatic Exposure Control for a digital mammography system, in: Proc. SPIE Vol. 7961, 2011: pp. 79615J1–79615J9

Summary

Breast cancer is a major health concern and a leading cause of cancer mortality among women. This has triggered breast cancer screening programs in several countries. Mammography is an established method for such screening actions with a breast cancer sensitivity reported to be 83.5%. The performance of screening programmes could be further improved if the number of missed cancers and unnecessary recalls could be reduced. Missed cancers may be due to overlapping tissue obscuring the important features of malignancy. Digital breast tomosynthesis (DBT) and breast computer tomography (BCT) are newly introduced technologies that remove these overlying tissues to some degree and may therefore improve the detectability of cancer. Some potential benefits of DBT include improvement in screening sensitivity, better identification of lesion size, improvement in characterisation, and decrease in recall rates. The investigation of these new devices and the development or improvements of prototype systems are expensive. That gave rise to dedicated simulation approaches that can be either a full simulation approach, from X-ray tube to a virtual clinical image, or a partial simulation approach in which lesions are simulated in existing images. After successful validation studies, simulation approaches might contribute to the optimization of different parameters, both in image acquisition and evaluation, as they allow to generate a lot of test material and to control several conditions

The main aim of this thesis was to develop a methodology to optimize the performance of breast tomosynthesis. We designed and validated a partial simulation framework to simulate 3D lesions into DBT images. We then developed 3D models of breast lesions (microcalcification clusters and masses). Observer studies were conducted to validate the realism of the appearance of these models into 2D and DBT images. These validations were followed by some applications as shown in the following brief descriptions of each chapter.

Chapter 2: We designed and validated a partial simulation framework to simulate 3D objects into DBT images. The simulation started by identifying case specific data including the acquisition settings, the phantom or compressed breast thickness and material. Ray tracing was performed to calculate 2D projections (i.e. x-ray transmission templates) of the 3D object for the x-ray spectrum that had been used to acquire the background images. The 2D templates of the object were modified to include object sharpness, then inserted at calculated positions in the projection images of the background while taking scatter corrections into account. These hybrid projections were then

reconstructed. The validation had been conducted using geometrically defined inserts simulated into background images of phantoms and compared against images of the objects acquired under the same conditions. The contrasts of the simulated and real objects were compared for different datasets. No statistically significant difference between real and simulated objects was found (p-values > 0.05). This framework was ready to simulate also other objects (like lesions) into phantom or patient DBT images.

Chapter 3: We proposed a new method of building 3D models of microcalcification clusters and described the validation of their realistic appearance when simulated into 2D and DBT images. Twenty-three breast biopsy specimens of microcalcification clusters with malignant and benign characteristics were scanned and their 3D reconstructed datasets were segmented to obtain 3D models of microcalcification clusters. The framework in chapter 2 was used to simulate forty microcalcification clusters into 2D and DBT projection images. The processed and reconstructed sets were shown to six radiologists in two separate observer studies. Results showed that the observers could not distinguish between the real and simulated clusters. This database of 40 clusters is suitable for use in future observer performance studies or as part of software breast phantoms.

Chapter 4: We worked at a new method of building 3D mass models with different morphological shapes and described the validation of the realism of their appearance after simulation into 2D and DBT images. Twenty-five contrast enhanced breast lesions were collected from MRI data, segmented, triangularly meshed and scaled to different sizes. These masses were referred to as non-spiculated masses and were then used as nuclei onto which spicules were grown following an iterative branching algorithm to form a total of 30 spiculated masses. These 55 mass models were projected into 2D and DBT images by means of the simulation framework in chapter 2. The realism of the appearance of these mass models was assessed by five radiologists in two observer studies. The results showed that the observers scored highly the realism of the simulated masses in both studies, but better in the 2D study. This indicated that the simulation procedure can still be improved for the DBT cases. This database is suitable for use in future observer performance studies whether in virtual clinical trials or in patient images with simulated lesions.

Chapter 5: We presented our first application of the use of the simulation framework described in chapter 2 to study one specific aspect of DBT. As mentioned earlier, the topic of microcalcification detectability and characterization in DBT is still under investigation with no conclusive results. For DBT systems equipped with an x-ray tube that moves continuously during exposure, some influence

of the focal spot motion on detectability and diagnosis of microcalcifications can be expected. In this study, we investigated the influence of position above the table on microcalcification contrast and signal difference to noise ratio (SdNR) in tomosynthesis images, by means of the simulation framework. Microcalcifications, represented by spheres of calcium with 400 µm diameter, were simulated into projection images of homogeneous objects and into anatomical backgrounds. The influence of system sharpness was included via the modulation transfer function (MTF) model that included detector, focus size, tube motion and x-ray oblique entry components. Scatter corrections were also considered. Results show contrast and SdNR reductions for spheres at increasing heights above the detector in all datasets. These results highlight the necessity of further developments in both the MTF and scatter to improve microcalcification visualization in DBT.

Chapter 6: We presented our first investigation of the morphological assessment of microcalcification clusters in 2D versus DBT. We focus on the morphology of calcifications because the morphological description of the shape of microcalcifications is a determining factor for recalling the patient or not. Four radiologists participated in the study and have described the shapes of microcalcifications in 71 clusters in 2D images and DBT series that were read in separate blinded sessions. Thirty five cases were real clusters and thirty six were simulated clusters by means of the simulation framework as described in chapters 2 and 3. An agreement test based on the kappa statistic was applied to evaluate the consistency of each reader's evaluation in 2D and DBT. Results have shown that there is good agreement between the observers' evaluations in these two modalities. This preliminary study has shown that the morphology of microcalcification clusters does not differ substantially in 2D versus DBT.

Chapter 7: Some concerns had been raised during the presentation of the previous chapter about the use of simulated clusters for the morphological assessment and the lack of a known morphological descriptor. Therefore, in this chapter, we conducted a study to evaluate fifty one real microcalcification clusters in 2D when compared to DBT in terms of morphology and number of calcifications detected in each cluster. Six radiologists evaluated 51 microcalcification clusters in paired 2D FFDM and DBT images. The two modalities were read in separate sessions to reduce bias. For each modality and for each cluster, the observers were asked to identify all the Le Gal types (morphology) present in the cluster and to count the number of visible microcalcifications. The results for the morphology assessment, between 2D FFDM and DBT, analysing each Le Gal type separately, showed that no systematic difference was found. All observers counted more calcifications in 2D FFDM than in DBT.

Summary

Combining the results in chapters 5, 6, 7, we can conclude that the interpretation of microcalcifications in DBT should be subject to further investigation and training.

The methodologies presented in this thesis are not limited to the context of the thesis. Further applications can be investigated using the simulation framework, and the lesions. The simulation framework can be used to guide the design of phantoms and test other parameters within the limitation of the framework. The 3D models of microcalcifications are part of the development of computer aided diagnosis (CADx) algorithm in use for future BCT and relying on the 3D features of the microcalcifications. Both 3D models, the microcalcification clusters and masses, are used in current studies in 2D. The models are simulated into 2D images of patients to compare different image processing algorithms in a 2-view study and to optimize the AEC settings in a 2D system. A number of studies can be conducted in patient images with simulated lesions in terms of detectability of lesions to compare 2D and DBT, to compare different image processing algorithms, to compare different reconstruction algorithms, etc. The 3D models can also be used in virtual clinical trials by inclusion in breast software phantoms. At the moment that proper breast software phantoms with realistic appearance, different glandularities, different structural shapes are available, these lesions can be combined and an unlimited number of investigations can be done to optimize the clinical performance of 2D and DBT systems with any type of geometries, combination of geometry, etc.

Samenvatting

Borstkanker is een belangrijke oorzaak van kankersterfte bij vrouwen die de maatschappij niet onverschillig laat. Mammografie is een gevestigde methode voor borstkankerscreening met een gevoeligheid van 83,5 %. De efficiëntie van screening kan verder worden verbeterd als het aantal gemiste kankers en overbodige vervolgonderzoeken zouden worden verminderd. Gemiste kankers kunnen te wijten zijn aan overlappende weefsels die de belangrijkste kenmerken van maligniteit verhullen. Digitale borst tomosynthesis (DBT) en borst-CT (BCT) zijn nieuw geïntroduceerde technologieën die het weefsel tot op zekere hoogte in drie dimensies weergeven en die daarom de gevoeligheid van het onderzoek en de aantoonbaarheid van kanker verhogen. Potentiële voordelen van DBT omvatten een verbetering van de screeningsgevoeligheid, een betere identificatie van de letselgrootte en een verlaging van het aantal onnodige vervolgonderzoeken. Het onderzoek naar deze nieuwe toestellen en de ontwikkeling of verbetering van prototypesystemen is duur. Daarom wordt gebruik gemaakt van simulaties. Ofwel kan men de volledige opstelling (met patiënt en letsel) simuleren om tot een volledige virtuele klinische studie te komen, ofwel wordt er geopteerd voor partiële simulatie, waarbij letsels gesimuleerd worden in klinische beelden. Na een succesvolle validatie van de simulatieketen kan deze bijdragen aan de optimalisering van verschillende parameters voor zowel beeldacquisitie en evaluatie. De kracht van een simulatieketen bestaat erin dat er op een vlotte manier nauwkeurig gecontroleerde condities ontstaan met bv. een gekend aantal letsels en welbepaalde goedaardige of kwaadaardige kenmerken.

Het belangrijkste doel van dit proefschrift was om een methode te ontwikkelen om de prestaties van DBT te optimaliseren. We hebben geopteerd voor partiële simulatie en daarvoor een simulatiekader ontworpen en gevalideerd specifiek voor 3D letsels in DBT beelden. We ontwikkelden 3D modellen van borstletsels (microcalcificatie clusters en massa's). Waarnemer studies werden uitgevoerd om te meten hoe realistisch de gesimuleerde letsels zijn. Na validatie kon partiële simulatie dan toegepast worden voor een paar heel specifieke vergelijkende studies van 2D versus DBT.

Hoofdstuk 2: Wij ontwierpen en valideerden een partieel simulatie methode om 3D objecten te simuleren in DBT achtergrondbeelden van patiënten of testobjecten. De input van de simulatie bestaat uit de acquisitie instellingen, de dikte en het materiaal van het fantoom of gecomprimeerde borst. Ray tracing werd gebruikt om 2D X-stralen projecties van de theoretische 3D objecten te berekenen voor het x-stralen spectrum dat was gebruikt voor de achtergrondbeelden. In eerste fase wordt de ideale

projectie van het object berekend. De object scherpte van de 2D projecties van het object worden dan aangepast aan het systeem waarvoor de simulaties worden doorgevoerd en dan worden de projecties ingevoegd op berekende posities in de achtergrondbeelden, met inachtneming van de effecten van strooistraling. Deze hybride projecties werden vervolgens gereconstrueerd. De validatie werd uitgevoerd met behulp van geometrisch omschreven objecten gesimuleerd in fantoombeelden en vergeleken met gemeten beelden van reële objecten. Het contrast van de gesimuleerde en de echte objecten werd vergeleken voor verschillende datasets en voor eenvoudige geometrische objecten. Er werd geen statistisch significant verschil gevonden tussen werkelijke en gesimuleerde objecten ($p > 0,05$). Dit was de eerste stap naar een volledige aanpak voor virtuele klinische studies gebaseerd op partiële simulatie.

Hoofdstuk 3: We stelden een nieuwe manier voor om 3D modellen van microcalcificatie clusters te maken en beschreven de validatie van hun realistische weergave na simulatie in 2D en DBT beelden. Drieëntwintig borstbiopsie exemplaren van microcalcificatie clusters met kwaadaardige en goedaardige eigenschappen werden gescand met een micro-CT en hun 3D gereconstrueerde datasets werden gesegmenteerd om 3D modellen van microcalcificatie clusters te verkrijgen. Het kader in hoofdstuk 2 werd gebruikt om veertig microcalcificatie clusters te simuleren in 2D en DBT projectie beelden. Zes radiologen hebben de verwerkte en gereconstrueerd beelden geëvalueerd in twee aparte studies. Resultaten toonden aan dat de radiologen de echte en gesimuleerde clusters niet konden onderscheiden. Deze database van veertig clusters is daarom geschikt voor gebruik in toekomstige observatie experimenten of als onderdeel van software borstfantomen.

Hoofdstuk 4: We stelden een nieuwe manier voor om 3D modellen te ontwikkelen van zachte weefsel massa's in de borst en voor de verschillende vormen waarin die beschreven zijn. We testten ook de validatie van het realisme van hun uitzicht na simulatie in 2D en DBT beelden. Vijfentwintig contrast versterkte MRI borstletsels werden verzameld, gesegmenteerd, wiskundig omschreven met driehoekig mazen en geschaald naar verschillende maten. Deze massa's werden aangeduid als onvertakte "non-spiculated" massa's en werden vervolgens gebruikt als kernen waarop tentakels (spicules) werden gegenereerd volgens een iteratief vertakkingsalgoritme om in totaal 30 vertakte massa's te vormen. Deze 55 massa modellen werden in 2D en DBT beelden geprojecteerd door middel van het simulatie kader in hoofdstuk 2. Het realisme van de verschijning van deze massa modellen werd beoordeeld door vijf radiologen in twee studies. De resultaten toonden aan dat de radiologen het realisme van de gesimuleerde massa even hoog inschatten als de echte massa's in de 2D studie, terwijl de scores voor DBT aantoonden dat de simulatie procedure weliswaar reeds bevredigend is maar nog zou kunnen

worden verbeterd. Deze database is geschikt voor gebruik in toekomstige studies van virtuele klinische proeven of medische beelden met gesimuleerde letsels.

Hoofdstuk 5: We hebben een eerste toepassing uitgewerkt die gebruik maakt van het simulatie kader in hoofdstuk 2. Zoals eerder vermeld zou DBT kunnen zorgen voor een betere karakterisering en detecteerbaarheid van letsels. Of dit ook het geval is voor microcalcificaties en voor alle DBT systemen moet nog onderzocht worden. Sommige DBT systemen zijn uitgerust met een röntgenbuis die continu beweegt tijdens de acquisitie. Dit kan bewegingsonscherpte veroorzaken en dus een invloed hebben op de diagnose van microcalcificaties. In deze studie onderzochten we de invloed van de verticale positie van de microcalcificaties op het contrast en op de verhouding van 'signaal-achtergrond tot ruis' (SdNR) in tomosynthesis afbeeldingen. Microcalcificaties, weergegeven door bollen van calcium met 400 micrometer diameter, werden gesimuleerd in projectiebeelden van homogene objecten en in anatomische achtergronden. De invloed van de beweging van de focus werd getest via de modulatie transferfunctie (MTF) model die detector, buis beweging en de schuine ingang van de x-straling inrichting incorporeert. Ook effecten van strooistraling werden gesimuleerd. Resultaten tonen aan dat contrast en SdNR voor bollen verminderden op grotere hoogte boven de detector voor alle datasets. Deze resultaten benadrukken de noodzaak van verdere ontwikkelingen in zowel de MTF en strooistraling onderdrukking om de weergave van microcalcificaties in DBT te verbeteren.

Hoofdstuk 6: We voerden ook een eerste onderzoek uit naar de morfologische beoordeling van microcalcificatie clusters in 2D versus DBT. We richtten ons op de morfologie van verkalkingen omdat de morfologische beschrijving van de vorm van microcalcificaties een bepalende rol is voor het terugroepen van de patiënt bij bevolkingsonderzoek naar borstkanker. Vier radiologen namen deel aan het onderzoek en hebben de vormen van de microcalcificaties beschreven voor eenenzeventig clusters in 2D en DBT beelden in verschillende blinde lezingen. Vijfendertig gevallen waren echte clusters en zesendertig werden gesimuleerd door het simulatie kader zoals beschreven in de hoofdstukken 2 en 3. Een overeenkomsttest op basis van de kappa statistiek werd toegepast om de consistentie van de evaluatie van elke lezer in 2D versus DBT te evalueren. De resultaten hebben aangetoond dat er een goede overeenkomst was tussen de evaluaties van de lezers voor deze twee modaliteiten. Deze inleidende studie heeft aangetoond dat de morfologie van microcalcificatie clusters niet wezenlijk verschilt in 2D versus DBT .

Samenvatting

Hoofdstuk 7: Een beperking van de vorige studie betreft het feit dat er geen pathologisch bilan ter beschikking is voor gesimuleerde letsels. Daarom hebben we in dit hoofdstuk een studie met eenenvijftig echte microcalcificatie clusters geëvalueerd in 2D en vergeleken met DBT qua morfologie en qua aantal verkalkingen, gedetecteerd in elke cluster. Zes radiologen beoordeelden eenenvijftig microcalcificatie clusters in 2D en DBT beelden. De twee modaliteiten werden gelezen in afzonderlijke sessies om vertekening te verminderen. Voor elke modaliteit en per cluster, werd de lezers gevraagd om alle Le Gal types (morfologie) in de cluster te identificeren en het aantal zichtbare microcalcificaties te tellen. De resultaten voor de vergelijking van de morfologie tussen 2D en DBT op basis van de Le Gal types afzonderlijk, toonde aan dat de resultaten lezer-afhankelijk zijn. Alle lezers telden meer verkalkingen in 2D dan DBT.

Als we de resultaten in de hoofdstukken 5, 6 en 7 combineren, kunnen we besluiten dat de interpretatie van microcalcificaties in DBT verder moet worden onderzocht.

De methoden in dit proefschrift zijn niet beperkt tot de context van het proefschrift. Verdere toepassingen kunnen worden onderzocht met behulp van het simulatie kader en de 3D letsels. Het simulatie kader kan worden gebruikt om het ontwerp van fantomen te leiden en andere factoren te optimaliseren Beide 3D modellen, de microcalcificatie clusters en massa's, worden vandaag al gebruikt in nieuwe studies, ook door andere groepen. De 3D modellen van microcalcificaties zijn onderdeel van de ontwikkeling van computer ondersteunde diagnose (CADx) algoritmes voor BCT die gebaseerd zijn op de 3D eigenschappen van de microcalcificaties. De modellen kunnen worden gesimuleerd in 2D beelden van patiënten om verschillende algoritmen voor 2D beeldverwerking te vergelijken en om de AEC instellingen in een 2D systeem te optimaliseren. Een aantal studies kunnen worden uitgevoerd in medische beelden met gesimuleerde letsels om de detecteerbaarheid van die letsels te vergelijken in 2D en DBT, om verschillende beeldverwerkingsalgoritmen te vergelijken, om verschillende reconstructiealgoritmen te vergelijken… De 3D modellen kunnen ook worden gebruikt in virtuele klinische proeven door ze te integreren in software fantomen van de borst. Op het moment dat er software fantomen zijn met realistische uiterlijk, verschillende densiteiten en verschillende structuren, kunnen deze letsels worden gecombineerd. Er kan dan een nieuw spoor ontwikkeld worden van onbeperkt onderzoek naar de klinische prestaties van 2D en DBT aan de hand van virtuele klinische studies. In de verdere fase volgt dan nog slechts de vraag of computer modellen de detecteerbaarheid in de klinische praktijk kunnen voorspellen

Professional Career Eman Shaheen

Eman Shaheen was born in Giza, Egypt in July 1982. She received her Bachelor's degree from the faculty of Computer Sciences and Information Technology, Cairo University in 2003. She received her Master degree in Video Processing in 2007. During the period of 2003 till 2007, she worked as teaching assistant in the faculty of Computer Sciences and Information Technology- Cairo University teaching undergraduate students image processing, fuzzy logic, pattern recognition algorithms. From July 2008, she joined the Medical Physics group in Radiology, University Hospital, KU Leuven, Belgium as a pre-doctoral student and co-worker in the European FP7 project Breast-CT. Since 2010, she started her doctoral training program under the supervision of Prof. Hilde Bosmans and Prof. Chantal Van Ongeval. The aim of her research project was to develop a methodology to optimize the performance of breast tomosynthesis by means of simulations.

International peer reviewed articles

1. **Shaheen E**, De Keyzer F, Bosmans H, Young KC, Dance DR, Van Ongeval C. "The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis" Medical Physics, accepted with minor revisions, 2014.
2. **Shaheen E,** Van Ongeval C, Zanca F, Cockmartin L, Marshall N, Jacobs J, Young KC, Dance DR, Bosmans H. "The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis". Medical Physics, 38(12), 6659-6671, 2011.
3. Zhang G, Pauwels R, Marshall N, **Shaheen E**, Bosmans H, Jacobs R. "Development and validation of a hybrid simulation technique for cone beam CT: Application to an oral imaging system". Physics in Medicine and Biology, 56(18), 5823-5843, 2011
4. **Shaheen E**, Zanca F, Sisini F, Zhang G, Jacobs J, Bosmans H. "Simulation of 3D objects into breast tomosynthesis images". Radiation Protection Dosimetry, 139(1-3), 108-112, 2010.

Refereed conference articles

1. **Shaheen E**, Bemelmans F, Van Ongeval C, De Keyzer F, Geeraert N, Bosmans H. "The investigation of different factors to optimize the simulation of 3D mass models in Breast Tomosynthesis" Proc. Of IWDM 2014.
2. Alquran H, **Shaheen E**, O'Connor JM, Mahd M. "Enhancement of 3D modelling and

classification of Microcalcification Clusters in Breast Computed Tomography (BCT)" Proc. of SPIE 2014.

3. **Shaheen E,** Marshall N, Bosmans H. "The influence of position within the breast on microcalcification detectability in continuous tube motion digital breast tomosynthesis" Proc. of SPIE 2013
4. Alquran H, **Shaheen E**, O'Connor JM, Mahd M. "3D Modelling and Classification of Microcalcifications in breast Computed Tomography (BCT)" MIPS 2013.
5. **Shaheen E**, Van Ongeval C, Zanca F, Cockmartin L, Marshall N, De Keyzer F, Young KC, Dance DR, Bosmans H. "The morphology of microcalcifications in 2D mammography and breast tomosynthesis: is it different?" Proc. of IWDM, 2012.
6. Elangovan P, Mackenzie A, Diaz O, Rashidnasab A, Dance D, Young KC, Warren L, **Shaheen E**, Bosmans H, Bakic P, Wells K. "A Modelling Framework for Evaluation of 2D Mammography and Breast Tomosynthesis systems." Proc. of IWDM, 2012
7. **Shaheen E,** Marshall N, Bosmans H. "Investigation of the effect of tube motion in breast tomosynthesis: Continuous or step and shoot?" Proc. of SPIE, 2011.
8. **Shaheen E,** Van Ongeval C, Zanca F, Cockmartin L, Marshall N, Bosmans H. et al. "Realistic simulation of microcalcifications in breast tomosynthesis". Proc. of IWDM, 2010.
9. Bosmans H, Jacobs J, **Shaheen E**, Zanca F, Cockmartin L, Bliznakova K, Lemmens K, Vollmar S, Kalender W. "Performance assessment of breast tomosynthesis systems: concepts for two types of phantoms". Proc. of IWDM, 2010.
10. Zanca F, Zhang G, Marshall N, **Shaheen E**, Salvagnini E, Marchal G, Bosmans H. "Software framework for simulating clusters of microcalcifications in digital mammography". Proc. of IWDM, 2010.
11. Zanca F, Bosmans H, Jacobs J, Michielsen K, Sisini F, Nens J, Young KC, **Shaheen E**, Jacobs A, Marchal G. "Contrast-Detail comparison between unprocessed and processed CDMAM Images." Proc. of SPIE, 2009.

Congress presentations

1. **Shaheen E.** et al. "The comparison between 2D digital mammography and digital breast tomosynthesis for morphological assessment of microcalcification clusters: a simulation study." RSNA, (Oral presentation) 2013.
2. **Shaheen E.** et al. "The development of a methodology to simulate 3D models of benign and malignant breast masses." RSNA, (Oral presentation) 2013.

3. **Shaheen E**. et al. “The morphology of microcalcifications in 2D mammography and breast tomosynthesis: is it different?” IWDM, (Oral presentation) 2012.
4. **Shaheen E.** et al. “Investigation of the effect of tube motion in breast tomosynthesis: Continuous or step and shoot?” SPIE, (Oral presentation) 2011.
5. **Shaheen E**. et al. “Creation of a database of 3D models of microcalcification clusters for future observer studies”. ECR, (Oral presentation) 2011.
6. **Shaheen E**. et al. “Towards a method for simulating lesions into digital breast tomosynthesis images”. ECR, (Poster presentation) 2010.

Submitted articles

1. **Shaheen E.** et. al. “The evaluation of microcalcification clusters in 2D digital mammography and breast tomosynthesis in terms of morphology and number of calcifications in the cluster”. (submitted to British Journal of Radiology).